LABOUR MARKET SEGMENTATION

IN THE NEW ENGLAND REGION

JOSEPH ANDREW STANISLAW

Ph.D. University of Edinburgh 1975

I declare that the work in this thesis has been composed by myself.

Acknowledgements

In writing this thesis my sincere thanks are due to my two supervisors, Professor J.N. Wolfe and Mr. Gavin C. Reid, who have unselfishly given me the benefit of their time, advice, and encouragement on many matters on this thesis. I am indebted to Mr. K.E. Anderson, formerly a Research Fellow in the Department of Geography, for his most valuable advice on the application of the techniques of classification analysis. I have also benefitted from suggestions and assistance made by Mr. Toby Morris, Programmer at the Edinburgh Regional Computer Center, on matters of data processing and computer programming. I also owe thanks to the members of the Boroughmuir Basketball Club, in particular, Ms. Andrea McInnes, Ms. Fiona Woods, Mr. Ken Johnston, Mr. Alf Bissett, and Mr. Ken Walker, for their assistance in typing, illustrating, and xeroxing this thesis. My special thanks are due to Ms. Gill Prentice for her expert labours on typing the final version of this thesis. I am greatly indebted to the University of Edinburgh for having provided me with an University Postgraduate Studentship that made my research possible. Finally, even though I already owe so much and am thankful to Gussie, my wife, in every aspect of my life, I wish to record my deepest thanks for her sympathetic interest and unending, patient encouragement throughout the preparation of this thesis. Needless to say, I alone am responsible for the work and views contained in this thesis.

J.A.S.

Edinburgh, July 1975

Abstract

The purpose in this thesis is to discuss the framework of labour market segmentation and provide a thorough and detailed description of the groups that are distinguished by labour market segmentation. As described by M.J. Piore, the concept of labour market segmentation depicts a labour market that is comprised of a set of three sub-markets or segments distinguished by different labour market characteristics and rules. A conspectus of the literature on the distribution of personal income and the labour market is given to outline the nature and direction of academic economic analysis that led to the development of this concept of labour market segmentation. Before an analysis of the labour market segments is conducted, it is necessary to have data that accommodate such an analysis. To this end, we transformed the 1970 New England Region Public Use Sample Data into a form suitable for our analysis. The data is used in a classification analysis of labour market segmentation. A statistical and illustrative presentation of the techniques of hierarchical classification and discriminant analysis is given. These are the techniques used in the empirical work for developing and analysing the labour market segments in terms of sixty socio-economic factors. The analysis serves to identify the socio-economic factors that distinguish the similarities and differences between the characteristics of the segments in the labour market.

Table of Contents

Table of Contents

CHAPTER 1

Introduction

1.1 Background

A primary consideration of economics is how to increase the degree to which all people share in economic growth and progress. In a sense, the income of a person can be used as a limited measure of equity and economic welfare. Income and, in turn, the problems of equity and welfare are the result of two economic processes, namely, production and distribution. Production and distribution are two interdependent processes. Income is created through production and distributed as a result of the relations of production. Thus, income is relevant not only to the problems of equity and welfare, but also to the problems encountered by the relations between the buyers and sellers that participate in the production and distribution processes. Both production and distribution occur within a market structure that determines the manner in which a person acquires income. In this thesis we are concerned with the underlying factors in such a market that influence an individual's position and ability to earn income in the market. As Lebergott (1959, p.335) wrote:

> We are concerned here with income as a measure of productivity rather than welfare, hence the ability of individuals to earn income under existing social and economic conditions. Conditions on the supply side have developed their talents for work, their abilities to invest, and have fired their desire for material goods. On the demand side still other factors have determined what places are available in the labour force, what skills and investments are requisite. Out of the interactions in the labour market come the rates of pay offered to ability and capital.

In the United States the distribution of personal income is a positively skewed distribution, that is, a relatively few people earn very high incomes and many people earn middle or lower incomes. Traditional economic analysis would presume that the incomes generating this distribution are created in a perfectly competitive labour market. In other words, the observed distribution is the result of an earnings process that maximises output from given resources and patterns of ownership and creates and distributes income to individuals according to the marginal productivity of the individuals. However, earnings and distribution are not sufficiently explained by the perfectly competitive labour market. There are many imperfections, such as employer associations, trade unions, governments, that might interfere with the competitive operations of the labour market. There are additional considerations besides marginal productivity that influence the determination of wages. For example, there are instances in which women and black men performing the same job as white men are paid less than white men. But, the perfectly competitive labour market that relies on marginal productivity to explain the allocation and pricing of labour is not completely adequate to accommodate these other aspects that influence the manner in which individuals earn income. More information is needed on the factors that influence the structure and operations in the labour market.

Therefore, alternative models of the labour market have been put forward. These alternatives, such as the two queue model, the dual labour market, and labour market segmentation, seek to incorporate the influences of non-market and non-competitive, as well as competitive economic factors into the structure and operations of the labour market. These alternatives view the structure of the labour market as the result of historical, political, and economic processes and changes that determine the conditions under which incomes accrue to individuals and groups. According to this view, income creation and distribution must be studied in terms of motivational patterns of groups in society within an institutional and economic framework resulting from historical, political, and economic processes. These alternative theories emphasise institutional changes and the points at which such changes occur. The alternative models, in particular labour market segmentation, analyse the processes that distinguish labour market differences and form groups of individuals and jobs in the labour market. These groups have been described in general terms, but have not been specifically identified in detail.

In the past, studies have examined groups in the earnings process, but not according to the descriptions offered in the alternative models of the labour market. The groups were arranged by income size, or industrial sector, or occupation by income size. But, shifts in thinking on the labour market structure and the development of new techniques of analysis now make it possible for us to examine the labour market in terms of groups of more or less homogeneous labour. The statistical analysis of such groups is not staightforward. The statistical techniques may serve to measure the results of accommodating the data within a particular framework, but do not pretend to reveal all the underlying motivations of the groups or to establish whether the framework itself is the most suitable. Neverthelass,the statistical techniques may provide results that may be an improvement over verbal descriptions. Thus, these techniques may serve to provide a means to establish a basic farmework for analysis of the groups distinguished by the alternative views of the labour market.

1.2 Purpose

The concept of labour market segmentation provides a possible workable framework of the labour market.[1] Labour market segmentation is the result of historical, political, and economic processes that have created and distinguished different segments with different types of individuals and jobs. The labour market is a set of sub-markets or segments distinguished by different labour market chracteristics and rules. The real-world imperfections of employer associations and trade unions and the non-market and non-competitive considerations of hiring policies and wage determination are taken into consideration in describing the labour market. The concept of labour market segmentation can indicate the position an individual may occupy in the labour market which in turn influences the individual's ability to earn income.

Most of the work on the concept of labour market segmentation has been of a theoretical or explanatory nature. This work has emphasised two points: firstly, the conditions that determine an individual's scope for action within the labour market; and secondly, the nature of the historical, political, and economic processes that created labour market segmentation. However, less emphasis has been placed on statistical and empirical analysis for identifying the individuals and jobs that constitute the labour market segments.

Therefore, the purpose of this thesis is to describe the concept of labour market segmentation and provide an analysis of the groups that are distinguished by it. To this end, the work in this thesis develops in four directions. Firstly, the thesis surveys the relevant literature on the distribution of personal income and the labour market,

---

1) See Piore (1972), Doeringer and Piore (1971), Doeringer (1971), and Piore (1970).

in order to establish the groundwork that led to a consideration of the concept of labour market segmentation. Secondly, because the data available were not in a form suitable for accommodating an analysis of the groups of individuals and jobs in the segments, the thesis transforms the available data into a form suitable for an analysis of labour market segmentation. Thirdly, the thesis discusses the techniques of analysis, namely classification analysis and discriminant analysis, that were used for identifying the groups in the labour market segments. Fourthly, the thesis discusses factors that may distinguish the labour market segments and performs an analysis of these factors to identify the similarities and differences within and between the labour market segments.

1.3 Scope

Because the concept of labour market segmentation is general and abstract and not necessarily accompanied by any set of data, the scope in surveying the literature is limited to establishing a general understanding of the concept. To go beyond a general overview would exceed the limited purposes of this thesis. Therefore, the scope is limited to describing the conditions, structure, and operations of labour market segmentation. The discussion will not seek to explain the historical and political processes that led to the creation of labour market segments. Thus, the discussion of labour market segmentation serves to establish its plausibility as an alternative framework for labour market analysis.

The concept of labour market segmentation provides a framework within which to determine an individual's ability to earn income. It is this framework that provides the basis for identifying the dominant groups in the earnings process. The discussion of labour market segmentation describes these groups and their motivations. However, this discussion and the subsequent empirical work do not undertake to consider explicitly the problems of equity, welfare, or specific motivations. The analysis provides an empirical and statistical description of the groups and may possibly indicate underlying associations between the groups. The analysis is intended to offer a description of the segments that goes beyond the verbal description in the discussion of labour market segmentation.

The description provided by Piore (1972) identifies three segments distinguished by labour market segmentation. This does not imply that there are only three segments in the labour market. Furthermore, the analysis does not pretend that the segments are perfect or ideal. The data sources and methods of data collection did not generate specific labour market data that reflected the factors and conditions described in the concept of labour market segmentation. In setting up the empirical work, it was often impractical or even impossible to obtain and to quantify the relevant factors necessary. Therefore, the segments are merely a statistical representation made with limited data. In the analysis, the emphasis is on ordering the limited data in a manner to achieve a first approximation of the characteristics of the segments distinguished by labour market segmentation.

1.4 Contents

To conclude, the arrangement of this thesis will be as follows. In Chapter 2 we shall discuss the facts and theories of the distribution of personal income. In this survey, we shall attempt to determine how successful the theories have been in explaining the distribution of personal income. Some of the theories are based on statistical approaches, whilst others adopt a more economic approach. The survey will help us to identify the important considerations and forces that underlie distribution and to indicate the general direction economic analysis has followed in attempting to explain the distribution of personal income.

From this discussion it will become clear that the labour market structure influences and contributes to the acquisition of personal income and the formation of the distribution. The labour market forms the subject of discussion of Chapter 3. We shall discuss the earnings process within the context of the labour market. In doing so, we shall present a discussion of the conventional approach to the labour market. Here we show that the conventional approach does accommodate many required revisions due to the introduction of more real-world factors. However, as more real-world factors are introduced, the assumptions and the adaptability of the conventional approach are strained. It is necessary to examine the alternative approaches to labour market analysis for the purpose of identifying an alternative workable model of the labour market. The alternative approaches develop socio-economic frameworks of the labour market that accommodate the recognition of non-market and non-competitive, as well as economic, factors. The alternatives hypothesise the existence of alternative models of the labour market, but do not specifically identify a workable, practical, and testable model.

But before any consideration can be given to whether the alternative models of the labour market are more appropriate than the conventional, it is necessary to put the alternatives into a systematic and workable framework. Chapter 4 is devoted to three aspects of this task. Firstly, we shall put forward the case for adopting the concept of labour market segmentation and specifying the groups of individuals (or segments). Secondly, we shall discuss the technical considerations of the problems of the availability of suitable data and the possibility of specifying perfectly homogeneous labour market segments. Thirdly, we shall examine the considerations involved in forming labour market segments according to occupational groupings.

In Chapter 5 we shall discuss the data used in this thesis. Firstly, we shall consider the type of data required in the analysis. Secondly, we shall discuss the data source in terms of the sample and collection procedure used to create the data. Thirdly, since the data was not in a form suitable for the empirical work in this thesis, we shall present the steps and methods involved in preparing the data for our analysis. Finally, we shal define the terms and concepts that relate to the data.

To carry out the necessary task of identifying the labour market segments, we must select the appropriate statistical techniques and procedures. This forms the context of Chapter 6. Here we present the techniques of analysis that are used in the empirical work of this thesis. In this chapter we discuss, in turn, hierarchical classification analysis and discriminant analysis. In doing so, we present the statistical framework, describe the method of interpreting the results, and provide an example of each technique of analysis. Each technique will contribute to an empirical and statistical description of the factors and associations underlying the labour market segements.

From this discussion we proceed to construct a realistic and relevant quantitative framework with which to identify and analyse the labour market segments. Chapter 7 is devoted to establishing this framework of analysis. Firstly, we shall discuss the general considerations involved in the use of our techniques of analysis for specifying and analysing the segments. Secondly, we shall present the procedure of analysis. Thirdly, the many factors responsible for the formation of the segments are discussed. In this discussion a priori expectations about the influence of the explanatory factors are presented. We have emphasised the factors that determine the supply of individuals to occupations and, in turn, to the labour market segments. Finally, as discriminant analysis requires known groups from which to establish criteria of classification, we shall specify a preliminary classification of a sample of the occupations into the labour market segments.

With the preliminary classification determined, we then move on in Chapter 8 to consider a discriminant analysis of this classification. With the technique of discriminant analysis we shall establish and discuss the important factors determining the classification. We shall attempt to account for the factors influencing the similarities and differences of the labour market segments on the basis of the framework discussed in Chapter 7. As a result we shall classify all the occupations not included in the sample on the basis of these factors. Furthermore, we shall carry out other discriminant analyses, but permitting fewer factors in the analysis. From this, we shall determine whether similar groups can be determined, but with less information required. In addition, we shall present an hierarchical classification analysis approach. This is an empirical approach to determine groups without involving subjective judgement. With these results we shall then perform a discriminant analysis to identify the important factors and assess the quality of the classifications. We shall compare the composition and the characteristics of the segments resulting from both types of analysis to assess to what extent the factors themselves as a reflection of the labour market structure determine the segments and to what extent similar factors are involved in both approaches.

Chapter 9 is devoted to providing a brief summary of the work in this thesis. Here we also attempt to present some thoughts on the general conclusions and the limitations of the empirical work for the concept of labour market segmentation. Although no firm conclusions can be reached at this stage about the possibility of adapting labour market segmentation to more general economic theories, we attempt to suggest a few lines of approach.

CHAPTER 2

Distribution of Personal Income

2.1 Personal Income Shares

The purpose of this chapter is to introduce the facts and the theories of the distribution of personal income shares. The distribution of personal income shares may mean different things to different people. For some, since distribution involves large numbers of individuals, explanations are sought in terms of statistical regularities. For others, a more economic approach, in which the value of a unit of labour is determined by scarcity and other economic principles, is used for explaining the distribution of personal income shares.

The distribution of personal income, sometimes referred to as the size distribution of incomes, relates to individuals and their income. Generally, economic analysis of the distribution of personal income puts the manner of acquiring incomes into the background. Historically, the significance attached to personal income shares was the size of the income (how much) and the distribution of the sizes of incomes rather than the manner in which incomes were acquired (how). Economic analysis concentrated on collective groups of individuals in which regular patterns might appear. Thus, in the past the emphasis of economic analysis in this area was on finding explanations for any regular patterns discovered.

At first glance, we may regard the distribution of personal income as no more than a graduation of incomes of the rich, the better off, and the poor. However, the importance of establishing such graduations is to provide a framework for measuring and explaining the degree of inequality observed in any distribution. The degree of inequality itself signifies more than merely who does or who does not receive a larger personal income. According to Lampman (1954, pp. 252-53), an unequal distribution of personal income reflects the inequality of personal contribution to national product, the inequality of personal consumption of that product, and the inequality of power to control resources.

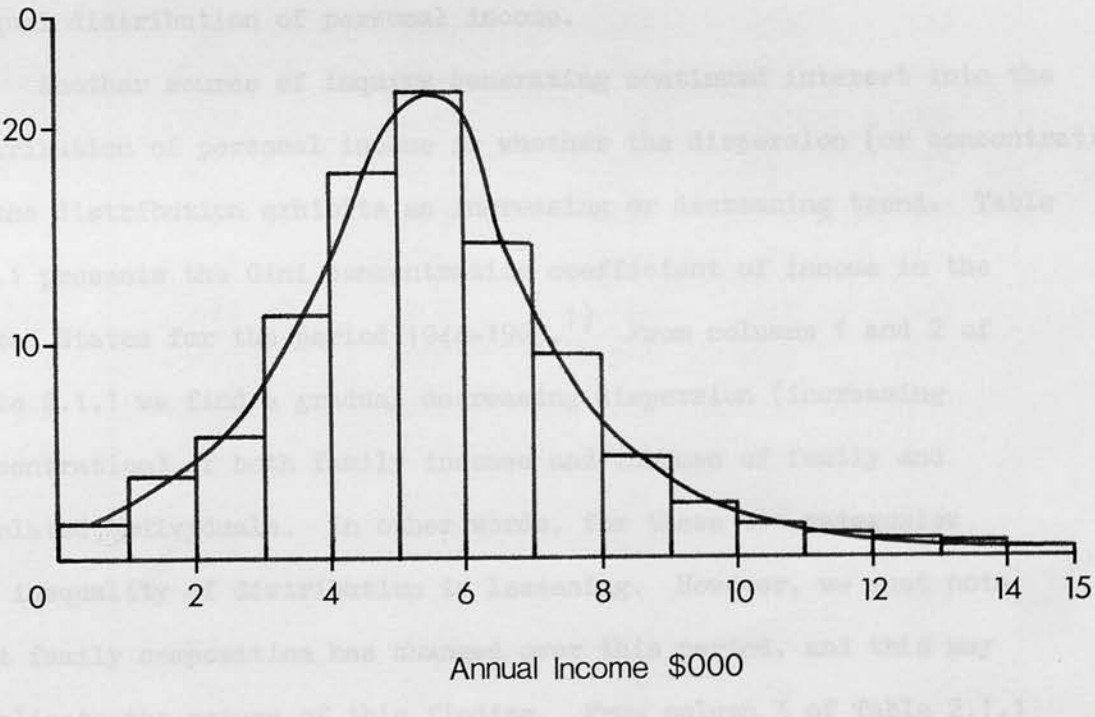
Furthermore, we may regard the distribution of personal income as reflecting the deferrment of income from current production to production in later periods. In this respect personal income measures the potential or actual command of individuals over economic goods. The control over economic goods bestowed by personal income, is in part a product of family status, total national income, individual tastes and preferences, and market structure. Therefore, we may consider the distribution of personal income as one imperfect measure of one aspect of economic welfare - the issue of equity. This issue of equity has often generated a confusion between seeking an explanation and passing moral judgment. However, the moral aspects of the distribution of personal income, such as judgments on equity, are beyond the limited purposes of this chapter.

The facts and the theories of the distribution of personal income shares may generate some misleading views of personal income. For example, it is generally believed that workers receive only wages income and capitalists only income from capital. This is misleading. Separating individuals on this basis ignores the facts of economic reality in which individuals, workers and capitalists, are often likely to receive both wages income and capital income.

We present the histogram in Figure 2.1.1 to provide an example

Figure 2.1.1

United States, 1959:
Frequency Distribution of White Male
Wage and Salary Workers aged 14 and over
Living in Central Cities of Urbanized Areas
who Worked 50-2 Weeks in the Year
by Wage and Salary Income
(Lydall, 1968, p.62)



of an observed distribution of personal income. In this example, the distribution is that of white males in the United States in 1959. We have presented this example because the shape of the distribution, the bold curve in Figure 2.1.1, depicts the shape of the typical distribution of personal income and family personal income observed in most Western economies. The most obvious feature of Figure 2.1.1 is that the distribution is positively skewed, that is, a large number of individuals receiving small or medium incomes and a few receiving large incomes. We would like to point out that this observation is at variance with the generally held belief that intelligence and ability are normally distributed and should be accordingly rewarded. It is this contradiction of beliefs that generated the fundamental inquiry into the nature of the unequal distribution of personal income.

Another source of inquiry generating continued interest into the distribution of personal income is whether the dispersion (or concentration) of the distribution exhibits an increasing or decreasing trend. Table 2.1.1 presents the Gini concentration coefficient of income in the United States for the period 1944-1965.[1] From columns 1 and 2 of Table 2.1.1 we find a gradual decreasing dispersion (increasing concentration) of both family incomes and incomes of family and unrelated individuals. In other words, for these two categories the inequality of distribution is lessening. However, we must note that family composition has changed over this period, and this may complicate the nature of this finding. From column 3 of Table 2.1.1

1) The Gini concentration coefficient of income is the area between the observed distribution of income (Lorenz curve) and the line of perfect equality dividied by the triangular area under the line of equality. It is calculated here by applying absolute weights to the differences in income between all pairs of observations standardised over the mean. The Lorenz curve plots the percentage of people against the percentage of total income that they receive.

Table 2.1.1

Income Inequality in the United States Among Families, Families and Unrelated Individuals, and For Persons 14 and Over with Income, by Sex: 1944-1965 (Schultz, 1969, pp. 78-9)

Year	Gini Concentration Coefficient of Income				
			Persons 14 and Over		
	Families (1)	Families and Unrelated Individuals (2)	Both Sexes (3)	Males (4)	Females (5)
1944	.4102	.4521	.505	.450	.525
1945	.3773	.4169	.487	.441	.496
1946*	-----	-----	----	----	----
1947	.3827	.4245	.483	.433	.517
1948	.3773	.4134	.474	.424	.511
1949	.3852	.4247	.489	.439	.522
1950	.3831	.4242	.498	.441	.535
1951	.3681	.4099	.480	.418	.521
1952	.3726	.4186	.470	.405	.511
1953	.3648	.4126	.484	.422	.528
1954	.3803	.4244	.489	.429	.519
1955	.3752	.4510	.509	.436	.542
1956	.3635	.4100	.506	.429	.532
1957	.3588	.4002	.506	.432	.534
1958	.3598	.4019	.503	.433	.543
1959	.3646	.4102	.515	.435	.546
1960	.3719	.4148	.517	.444	.531
1961	.3805	.4265	.527	.456	.551
1962	.3642	.4102	.514	.441	.541
1963	.3651	.4146	.516	.443	.536
1964	.3607	.4141	.511	.442	.537
1965	.3658	.4136	.512	.447	.535

we conclude that there is an increasing dispersion of personal incomes. If we examine personal income according to the sex of the recipient, (columns 4 and 5) we find slight trends toward an increasing dispersion of both male and female incomes. If we exclude the data for the year 1944, we find that there is no significant tendency for the equalisation of either family or personal incomes.

The distribution of personal income shares is the outcome of an intricate circular economic process of operations; production, distribution, and consumption. Income is created and distributed. Any attempt to explain or criticise the distribution of personal income is consequently an attempt to explain or criticise the underlying economic process or mechanism creating income. Therefore, we regard the survey of the theories and explanations of the distribution of personal income that follows as a necessary first step in clarifying, ordering, and criticising. This survey serves to illustrate the approaches and the factors that economic analysis generally considered in explaining the nature of distribution. In addition, it is this survey that will lead us from a discussion of the distribution of personal income (how much) in the economic system as a whole to a more fundamental discussion in the subsequent chapter on the earnings process and market structure (how). To go beyond a general conspectus would exceed the limited purposes of this chapter.

We shall adopt the terminology of Bjerke (1970, p. 235) for classifying and ordering the theories and explanations in the survey. Firstly, we will consider the "theoretical-statistical" approach. This approach views the distribution of personal income shares in terms of stochastic processes. The three basic methods have been the fitting of statistical frequency functions, the development of

Gibrat's "law of proportionate effects", and the specification of transition probabilities. Thus, in this approach, anonymous random forces are generating the distribution but are so many and so diverse that none may be separated from the mass influence. The second approach is the "sociological" approach. This approach attempts to identify, in addition to economic factors, a number of institutional factors which influence the income level of an individual and the distribution of personal income. Thus, in the sociological approach important social, institutional, and economic forces acting upon the distribution are preeminent.

2.2 Theoretical - Statistical Approach

These approaches take theoretical - statistical views of the formation of the distribution of personal income. In other words, it is believed that stochastic processes implying an indeterminate number of random forces generate the distribution. In one of these approaches one fits a statistical frequency function to the distribution of personal income. The distribution of personal income is generally regarded as a positively skewed distribution, as in Figures 2.1.1 above. The task then is to fit a statistical frequency function to a skewed distribution.

One particular statistical frequency function, the Pareto, has been fitted to the skewed distribution. The Pareto distribution was the result of studies by Pareto into apparent regularities in the curves of the distributions of personal income from various countries. In these studies, he constructed tables of frequency distributions and then illustrated them in diagrammatic form. From his observations Pareto found that there seemed to be a fixed relation between a given income level and the cumulated percentage of individuals earning this income and higher.

In simple form we can write this relation as the cumulative (more than) distribution function

$$P(y) = Ay^{-\alpha} \quad ; \qquad (2.2.1)$$

where $P(y)$ is the percentage of units in excess of y; y is the income level; and A and α are the parameters. In logarithmic form the cumulative (more than) distribution function (2.2.1) becomes

$$\log P(y) = \log A - \alpha \log y \quad . \qquad (2.2.2)$$

Thus, we find in the Pareto Law that the logarithm of the percentage of individuals earning a given income or higher is a negatively sloped linear function of the logarithm of that income level.

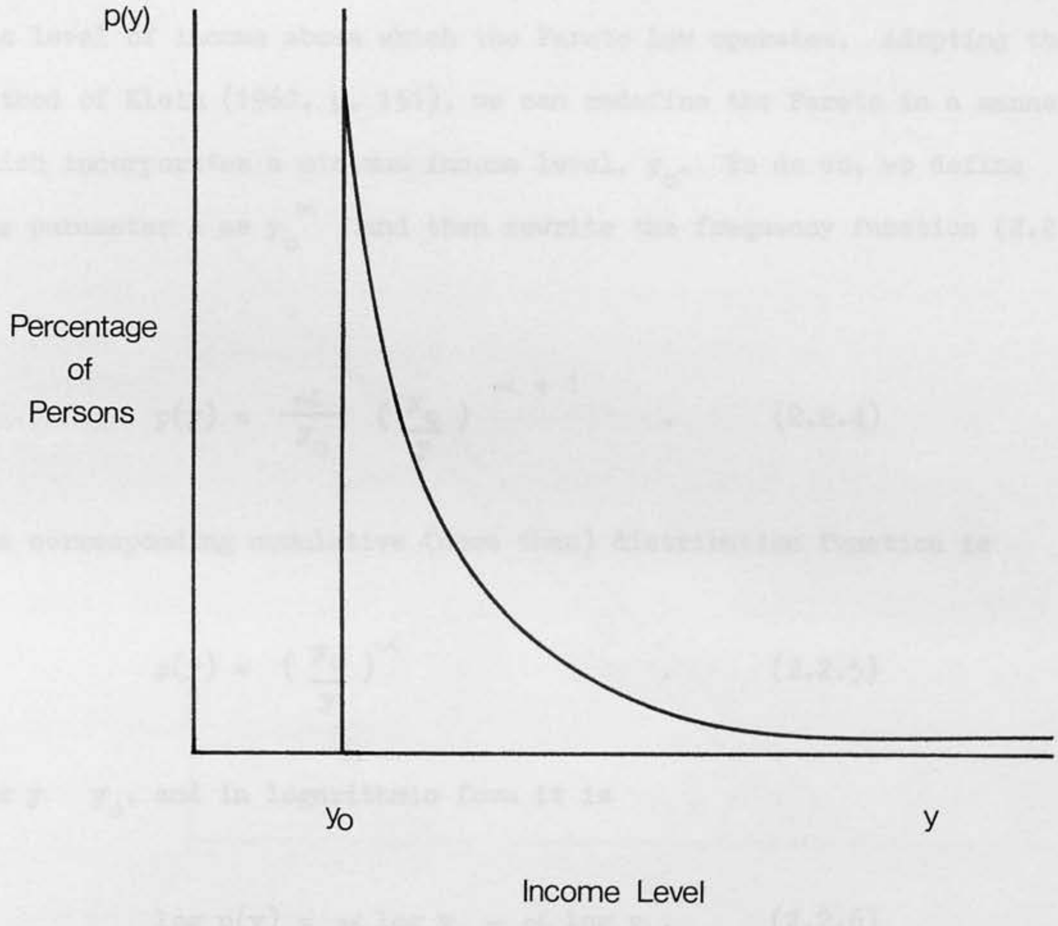
The corresponding frequency function derived from (2.2.1) is

$$p(y) = \alpha Ay^{-(\alpha+1)} \qquad (2.2.3)$$

where y is usually considered greater than some minimum level, y_o. A typical Pareto frequency distribution is presented in Figure 2.2.1. The income levels are marked along the horizontal-axis, and the percentage of persons with a given income is marked along the vertical-axis. From the frequency function (2.2.3) and the curve in Figure 2.2.1 we conclude that as income levels approach zero, the relative frequency approaches infinity. As income levels get progressively larger, the relative frequency falls toward zero. The typical Pareto frequency distribution in Figure 2.2.1 portrays the

Figure 2.2.1

Typical Pareto Frequency Curve



long tail to the right of the typical distribution of personal income, but does not depict the hump usually associated with the left side. For this reason the Pareto is considered most applicable to upper income levels.

In practice, the Pareto has been used to analyse levels of income above some minimum level, y_o in Figure 2.2.1. We consider this the level of income above which the Pareto Law operates. Adopting the method of Klein (1962, p. 151), we can redefine the Pareto in a manner which incorporates a minimum income level, y_o. To do so, we define the parameter A as y_o^{α} and then rewrite the frequency function (2.2.3) as

$$p(y) = \frac{\alpha}{y_o} \left(\frac{y_o}{y} \right)^{\alpha + 1} . \qquad (2.2.4)$$

The corresponding cumulative (more than) distribution function is

$$p(y) = \left(\frac{y_o}{y} \right)^{\alpha} . \qquad (2.2.5)$$

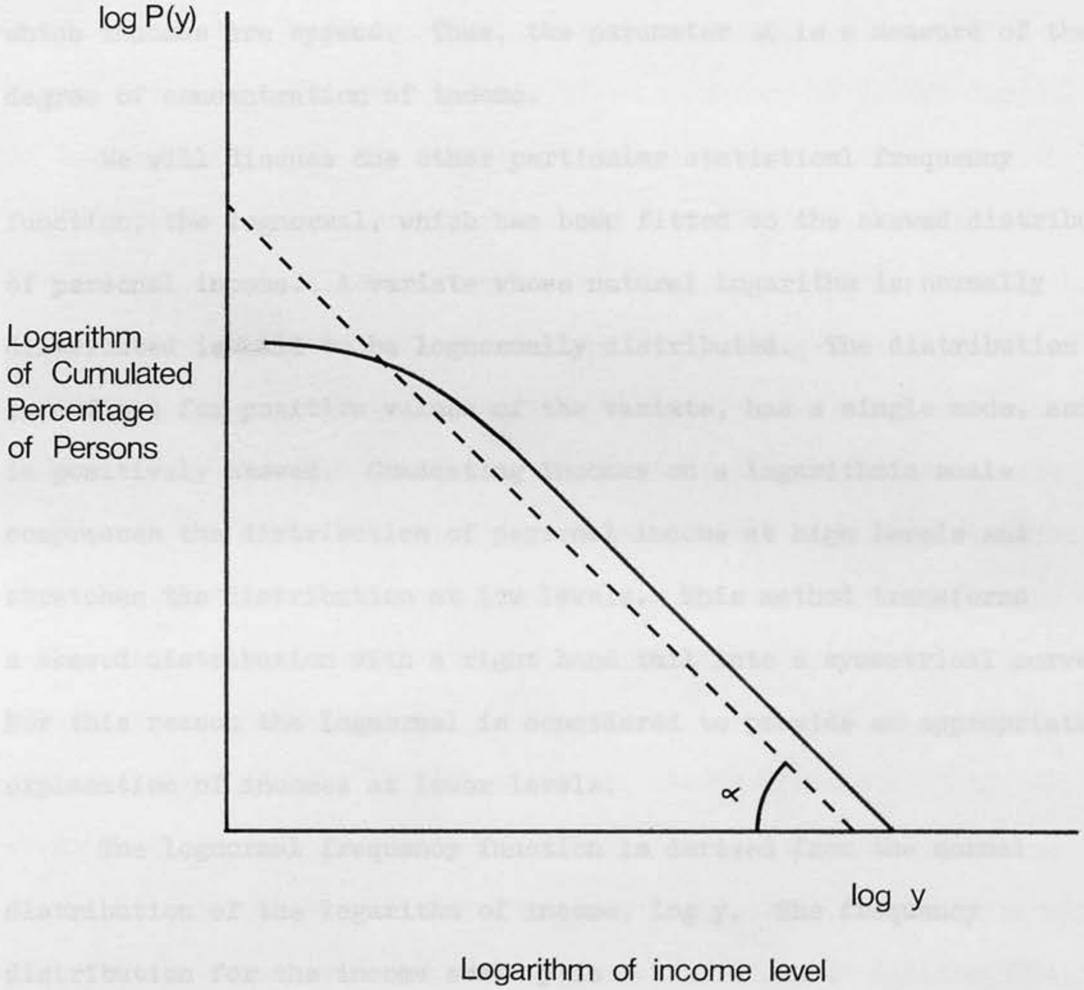
for $y \quad y_o$, and in logarithmic form it is

$$\log p(y) = \alpha \log y_o - \alpha \log y . \qquad (2.2.6)$$

A typical Pareto comulative (more than) distribution presented on a double logarithmic scale is illustrated in Figure 2.2.2. The dotted straight line represents a negatively sloped linear function, and the smooth line represents the typical curve obtained from data. As a rough guide, the more closely a scatter of points bunches about the dotted straight line, the more willing we are to assent that the

Figure 2.2.2

Typical Pareto Curve (Double Logarithmic Scale)



distribution is Paretian.

We can conclude from Figure 2.2.2 that α measures the slope of the Pareto curve to the log y-axis. The steeper or more vertical the line, the greater is the value of α, and the narrower the income range in which the group of incomes is concentrated. The more horizontal the line, the lower is the value of α, and the wider the range over which incomes are spread. Thus, the parameter α is a measure of the degree of concentration of income.

We will discuss one other particular statistical frequency function, the lognormal, which has been fitted to the skewed distribution of personal income. A variate whose natural logarithm is normally distributed is said to be lognormally distributed. The distribution is defined for positive values of the variate, has a single mode, and is positively skewed. Graduating incomes on a logarithmic scale compresses the distribution of personal income at high levels and stretches the distribution at low levels. This method transforms a skewed distribution with a right hand tail into a symmetrical curve. For this reason the lognormal is considered to provide an appropriate explanation of incomes at lower levels.

The lognormal frequency function is derived from the normal distribution of the logarithm of income, log y, The frequency distribution for the income scale y is

$$p(y) = \frac{1}{\sigma\, y \sqrt{2\pi}} \exp\left[- \frac{1}{2\sigma^2} (\log y - \mu)^2 \right] \tag{2.2.7}$$

The parameters of the distribution are μ, the mean of log y, and σ^2, the variance of log y.

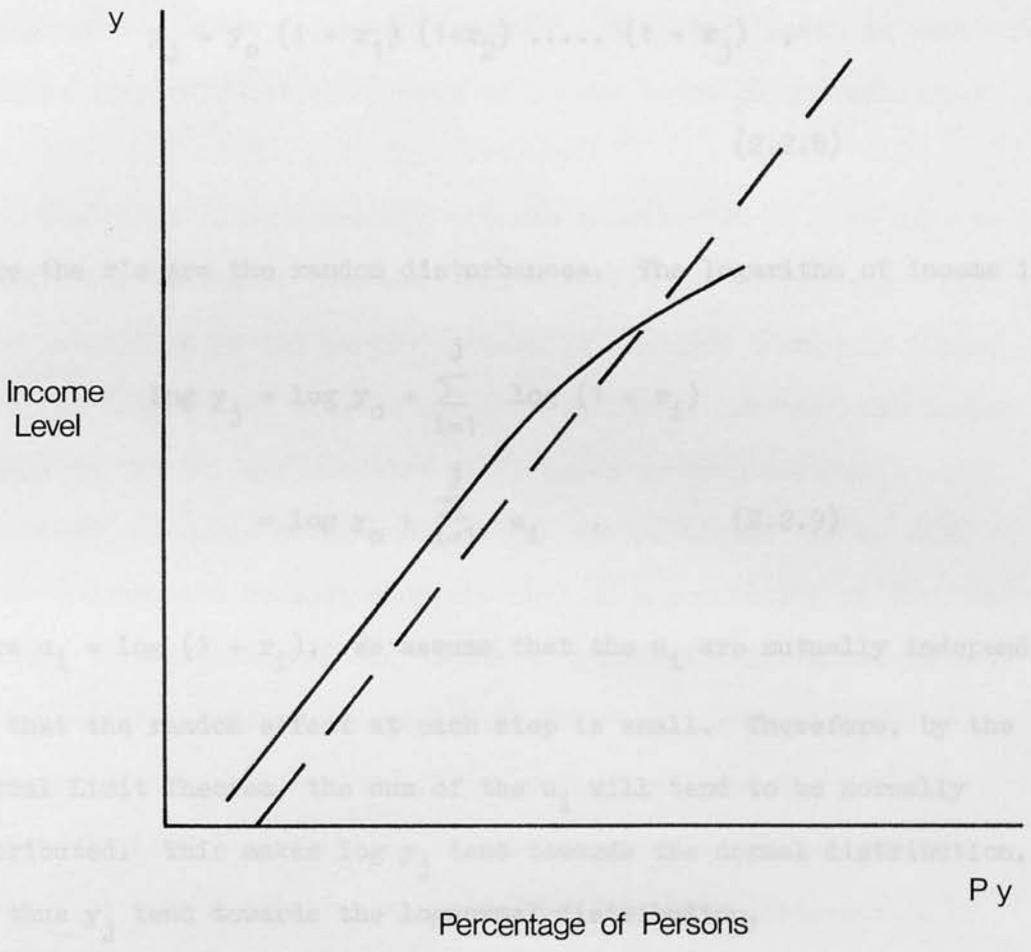
The typical lognormal cumulative distribution function plotted on a logarithmic scale is presented in Figure 2.2.3. We plot the incomes along the vertical-axis and the cumulated percentage of persons along the horizontal-axis. The dotted straight line represents the cumulative distribution, and the smooth line the typical curve from data. As in the Paretian, the linearity criterion applies. From Figure 2.2.3 we also conclude that σ is a measure of income inequality. In the lognormal then the degree of inequality varies directly with σ.

Even though statistical frequency distributions, such as the Pareto and lognormal, seem to graduate the positively skewed distribution of personal income, this fact is an empirical observation. To be of relevance, these observations must have some foundation in economic theory. For this reason, we now turn to a discussion of the theories that were devised to give rise to the statistical frequency distributions. These theories are based on assumptions about random processes that influence an individual's ability to raise his income level. These processes were introduced to explain the discrepancy between the normal distribution of abilities and the skewed distribution of income. Even though in much of standard statistical theory random processes generally generate normal distributions, in the case of the distribution of personal income random forces give rise to a skewed distribution.

The first of these processes that we will consider is the "law of proportionate effects" that gives rise to the lognormal distribution. The "law of proportionate effects", as defined in Aitchison and Brown (1969, p. 22), states that "the change in the variate at any step of the process is a random proportion of the previous value of the variate." Suppose that the initial value of the variate is the

Figure 2.2.3

Curve of the Typical Lognormal Cumulative Function



initial income, y_o, of an individual, and that after the jth step in the process of random proportional changes, $(1 + r_j)$, it is y_j. The random changes, $(1 + r_j)$, are mutually independent. The final value would be y_n after n steps in the process. At the jth step of the process the original income, y_o, will have become

$$y_j = y_o (1 + r_1)(1+r_2) \ldots\ldots (1 + r_j) \quad , \tag{2.2.8}$$

where the r's are the random disturbances. The logarithm of income is

$$\log y_j = \log y_o + \sum_{i=1}^{j} \log (1 + r_i)$$

$$= \log y_o + \sum_{i=1}^{j} u_i \quad , \tag{2.2.9}$$

where $u_i = \log (1 + r_i)$. We assume that the u_i are mutually independent and that the random effect at each step is small. Therefore, by the Central Limit Theorem, the sum of the u_i will tend to be normally distributed. This makes $\log y_j$ tend towards the normal distribution, and thus y_j tend towards the lognormal distribution.

The "law of proportionate effects" generated the lognormal distribution of personal income through a process of random factors acting multiplicatively. The process implies that random factors either increase or decrease proportionately the number of income recipients at an income level, but that the factors act randomly regardless of income level. From the observation that the lognormal distribution of personal income was generated by the "law of proportionate effects" we can further infer four underlying conditions about movements between income levels. The first is that the high income recipients and the low income recipients have the same average proportional growth in income level. The second is the dispersion of growth rates around the common average is the same for high and low income recipients. The third condition is that the distribution of proportionate growth rates is lognormal. And, the fourth is that the relative dispersion of the levels of income tends to increase over time.

The other random economic process constructed to give rise to a skewed distribution of personal income is the Markov process. The basic assumption of the Markov process is that any change in income during an interval of time is a random variable. Through the Markov process an initial distribution will, under certain regularity conditions, converge towards some final distribution. If we have income information on each economic unit of a population in two adjacent time periods, we can construct a matrix whose elements show the number of income receivers moving from one income class in the first period to another class in the subsequent period. The row totals constitute the frequency distribution of the first period, whereas the column totals constitute the frequency distribution of the subsequent period.

By dividing a typical row element of the matrix by its corresponding row sum in the first time period, we obtain the transition probabilities. These define the probability of moving from a given income class in the first period to another income class in the next period. The transition probabilities are assumed to be constant through time. We find that the matrix yields a set of linear simultaneous equations that describes how the distribution in the first period is transformed into the distribution in the subsequent period. With the system of linear simultaneous equations we can determine the expected numbers in each income class in future periods. The system can be shown to generate a limiting distribution. The limiting distribution depends entirely on the transition probabilities and is independent of the initial distribution. By varying the restrictions imposed on the Markov process, different distributions are generated.

An example of a specific theory constructed to explain the distribution of personal income in terms of random economic processes is Champernowne's (1953). Champernowne starts from the assumption that the distribution of income is Paretian and analyses this assumption in terms of the Markov process and the specification of transition probabilities. He stipulates five restrictions under which the random process will generate a Pareto distribution from a randomly distributed set of incomes. The five restrictions are:

1. The lengths of the income ranges are proportional, each successive income range being larger than the one beneath by the same proportion. That is, the income class ranges are in geometric progression.
2. No income receivers move up by more than one interval range in a year, or down by more than n income ranges in a year, for some fixed integer n.
3. The number of incomes is constant and the incomes live on, although their recipients are transitory.
4. The transition probabilities depend on the spread between two adjacent classes.
5. The "law of proportionate effects" applies to the transition probabilities.

With these restrictions he shows that the distribution of personal income will, in the limit, be distributed according to the Pareto Law.

One other example of this type of approach is that of Aitchison and Brown (1969). They examined the effects of modifying the restrictions imposed by Champernowne. They altered Restriction 1 by forming an arithmetic progression of income classes, and Restriction 4 by making the transition probabilities depend on the ratio of incomes in two adjacent periods. They show that these modifications will give rise to the lognormal distribution in the limit, although it is generated by a different mechanism from that described by the "law of proportionate effects".

The theoretical-statistical approaches discussed above have been prominent in economic analysis. But they suffer from two basic weaknesses. The first is that fitting a statistical frequency function to the distribution of personal income does not facilitate an understanding of the entire distribution. Thus, we find the Pareto only explaining the higher income portion of the distribution and the lognormal only explaining the lower and middle income portion of the distribution. The second weakness is that reliance is placed upon stochastic processes and that the distribution of personal income is inevitable. These processes consider the characteristics of the population as irrelevant to the final outcome. However, income today does not only depend on income in the previous period. Income depends on a number of factors, as we shall see in the subsequent sections of this chapter, and the influence of these factors in the random economic processes cannot be ignored.

It is at this point that we turn to two other theoretical-statistical approaches that not only rely upon the laws of probability and chance but also introduce characteristics of individuals into the random process generating the distribution of personal income. The first of these approaches is that of Friedman (1953). Basing the approach on probability and chance, he developed an explanation of the shape of the distribution curve in terms of individual choice. Friedman (1953, p. 278) assumed:

> The alternatives open to an individual differ, among other respects, in the probability distribution of income they promise. Hence his choice among them depends in part on his taste for risk. Let the same set of alternatives be available to members of two societies, one consisting of people who have a great aversion to risk; the other, of people to 'like' risk. This difference in tastes will dictate different choices from the same alternatives. Individuals choose among alternatives involving risk as if they knew the probability distribution of incomes attached to each alternative and were seeking to maximise the expected value of some quantity.

In fact, the differing propensities to undertake risk generate the distribution of incomes. Risk-liking persons will enter occupations or industries in which there are large and variable chances for profit and loss, while risk-averse persons will enter occupations with little fluctuation in profit and loss. On the one hand, the distribution of income of the risk-liking persons will be skewed because a few persons beat the risk factor and receive high rewards, while the losers receive moderate or low reward. On the other hand, the distribution of the risk-averse persons will be normal because the individuals receive income according to the distribution of abilities, a normal distribution. The overall distribution, the added total of the distributions from the two types of proprietry to risk, is a positively skewed distribution. There is neither a special reward for those persons taking a risk and losing nor for those playing it safe and undertaking no risk. These persons create the "hump" in the distribution of personal income. This explanation accounts for the characteristic tail at the high income levels and offers one of the few explanations in the theoretical-statistical approach for the absence of a tail at the lower income levels of the income distribution.

Individual choice as described by Friedman is only accurate to the extent that production risks are constantly undertaken by entrepreneurs. Pen (1971, p. 243) pointed out a subtle twist in this form of lottery. "It is striking that the same people keep on winning and others never do. In genuine lotteries this would attract the notice of the police. The super rich do not always run such big risks - they may spread their business interest over many small firms. Dangerous risks are often borne precisely by the small businessman, and he consequently often becomes the victim of the competitive process. Moreover, the worker's existence is not free from the chance of things going wrong either, whether he likes it or not." Furthermore, individual choice is constrained by an individual's characteristics and the market structure. Nevertheless, this approach established the need of recognising individual choice as part of the process of distribution.

However, other factors are at work in the distribution process. The last theoretical-statistical approach we will present, that of Lydall (1968), introduces more elements that may influence the distribution of income. Lydall is moving towards the sociological approach in that he considers factors such as the differences in ability, the influence of social status, the differences in education, and the structure of industrial organisation. To begin, he assumed that a person's position in the hierarchical structure is an unspecified function of ability, education, and social status. He then proceeded to present a process that fits the Pareto distribution to the distribution of personal income.

Lydall (1968, pp. 128-29) assumed that employees within an organisation are arranged in a pyramid formation according to hierarchical grades, G_i, where $i = 1, 2, \ldots k$ with 1 as the lowest grade. If y_i is the number of employees in grade i, G_i, then we can write his first assumption as

$$\frac{y_i}{y_{i+1}} = n, \tag{2.2.10}$$

where $n > 1$ and fixed for all i, that is, on average a person in a given grade controls a fixed number of people in the grade below. If x_i is the standard wage for G_i, we can write his second assumption as

$$\frac{x_{i+1}}{nx_i} = p, \tag{2.2.11}$$

where $p < 1$ and fixed for all i, in words, the income of a given grade is a fixed proportion of the total income of the persons in the grade below. He further assumes that $\frac{x_{i+1}}{x_i} > 1$ and $np < 1$.

If the top grade G_k has only one person, the number of persons in G_{k-1} is n, and G_{k-2}, is n^2, and hence

$$y_i = n^{k-1}, \tag{2.2.12}$$

The total number of employees in all grades above G_i is

$$Y_i = 1 + n + n^2 + \ldots + n^{k-1}$$

$$= \frac{n^{k-i} - 1}{n - 1} . \tag{2.2.13}$$

Thus, the proportion of the Y_i employees to all the employees in the firm can be written as

$$Q_i = \frac{Y_i}{Y} = \frac{n^{k - i + 1} - 1}{n^k - 1} \approx n^{1 - i} \qquad (2.2.14)$$

where n or k is reasonably large.

From (2.2.11) we have

$$x_i = (np)^{i - 1} x_1, \qquad (2.2.15)$$

from which we can obtain

$$i - 1 = \frac{\log x_i - \log x_1}{\log np} \quad . \qquad (2.2.16)$$

Substituting for (i - 1) in (2.2.14), we can rewrite (2.2.14) as

$$\log Q_i = \lambda \log x_1 - \lambda \log x_i, \qquad (2.2.17)$$

where $\lambda = \frac{\log {}^1/n}{\log np} > 1$, since $p < 1$ and $np > 1$. Expression (2.2.17) is the Pareto distribution. The measure of the Pareto coefficient is λ . This process suggests that the salary of the upper income levels will follow the Pareto distribution, and that the salary of a person in a given grade is a function of the number of persons in the grade below. This approach is appropriate for treating structured and bureaucratic organisations. However, it cannot handle, nor does it consider, the less organised sectors of the market. Even though ability, education, and social status were acknowledged in this approach, they were not sufficiently taken into account, nor were different institutional structures.

2.3 Sociological Approach

The theoretical-statistical approach tends to describe the facts of the distribution rather than to present an explanation of the facts. It subsumes a number of specific influences on the distribution of personal income under the category of random effects. By contrast, the sociological approach tends to illuminate the individual components influencing the distribution. It is assumed that determining factors other than ability intervene to distort the relation between ability (normal distribution) and earnings (positively skewed distribution). In the sociological approach the shape of the distribution curve depends on a number of sociological and institutional factors, such as age, sex, trade union influence, discrimination, geographical location, and others. To establish the basic differences in the approaches Bjerke (1970, pp. 242-43) wrote that "if changes occur in these factors and if these changes affect income levels, the shape of the income distribution will be changed. The theoretical-statistical approach assumes a process leading to convergence towards a definite type of distribution, whereas the sociological approach assumes that the shape of the income distribution at any given time is the result of a historical process, and the shape may therefore be changed when the relative importance of the different factors is changed."

The first sociological approach which we will review is the human capital approach. It concentrates upon the activities that influence the future monetary and psychic income of an individual. Human capital is the investment in an individual's resources, for example, education, training, improved health resources. Levels of investment in human capital are associated with positions on the distribution curve. Income is determined by supply and demand conditions for individuals possessing different types of human capital. The supply condition is determined by an individual's human capital and his energy to exploit his position which we will call "economic talent". Two individuals with the same investment in human capital, other factors equal, possess the same potential to obtain the economic rewards. But, they are not guaranteed the same economic reward because both do not necessarily possess the same "economic talent", that is exploit their position with the same vigour and energy.

The human capital approach has been developed by Becker (1964). Becker and Chiswick (1966), Mincer (1958), and others. It assumes a free choice of individuals to invest in the various forms of human capital. In other words, a kind of equality of opportunity as "a situation in which low parental wealth and other supply disadvantages were sufficiently offset so that the effective supply curve of funds was the same for everyone. One way to achieve this would be to make investment in human capital a free good through subsidies from public and private agencies; all supply curves, in effect, would then lie along the horizontal axis. Our definition of equality of opportunity would imply not equal investment but equal opportunity to invest, the actual amount depending on ability and other personal characteristics."

From the human capital approach we can infer that different levels of investment in human capital lead to the positive skewness in the distribution of personal income. Workers with higher levels of investment in human capital have more possibilities of applying and perfecting their knowledge and skills than do those with lower levels of investment in human capital. Even though two workers may have the same investment in human capital, one may seem more than the other. The one who earns more has employed greater "economic talent" in exploiting his human capital.

In the simple human capital model presented by Becker (1964, p. 62), earnings, Y, after an investment period are approximated by

$$Y = X + rC \quad . \tag{2.3.1}$$

C measures the total investment, r is the average rate of return on the investment, and X is the earnings when there is no investment in human capital. If X is ignored, the earnings depend on r when C is held constant. Earnings will depend on the average rate of return on the investment in human capital.

In developing the human capital approach it is further assumed that the more able individuals invest more in human capital than others. So, ability and investment in human capital should be positively correlated. Thus, if ability and earnings both varied, income would tend to be skewed. The more able individuals have an economic incentive to invest large amounts in human capital. This tends to provide a reconciliation of the positive by skewed distribution of personal income with the normal distribution of abilities.

To generalise equation (2.3.1) Becker (1964, p. 65) puts

$$Y_j = X_j + \sum_{i=0}^{j-1} r_i C_i + (-C_j), \tag{2.3.2}$$

where j refers to the current year and i to the previous years, C_i measures the investment cost at age i, C_j measures the current costs, and r_i refers to the rate of return on investment. X_j is the earnings in the current year attributed to the returns from the original level of investment in human capital. The smaller the proportion of earnings resulting from the investment in human capital, $\sum_{i=0}^{j-1} r_i C_i$, relative to X_j, the more distribution of earnings is dominated by X_j. The lower skill groups (small investments in human capital) have their earning power influenced by X_j; while the higher skill groups (larger investments in human capital) have their earnings dominated by $\sum_{i=0}^{j-1} r_i C_i$.

At earlier ages investment costs are high; therefore, $- C_j$ should reduce Y_j. And, X_j should be relatively more important than $\sum_{i=0}^{j-1} r_i C_i$, resulting in a less skewed distribution at younger ages. At older ages investment costs decrease each year, so $- C_j$ should have less impact. However, with increasing age the returns on previous investment $\sum_{i=0}^{j-1} r_i C_i$, should tend to increase in importance and to generate a positively skewed distribution of income.

The other sociological approaches to be discussed are not only concerned with explaining the positively skewed distribution of personal income, but also with seeking the causes of incomes being concentrated in the middle of the distribution. These approaches emphasise two trends within Western economies; firstly, the tendency towards increasing centralisation, and secondly, the growth of large-scale organisations. Two practical consequences of these trends were the spread of universal education and technological development. During the development of an advanced economy, such as the United States, industry demands exceeded the supply of labour with the necessary skills. On the one hand, this shortage encouraged the spread of universal education which, in turn, generated an increase in the numbers of skilled and semi-skilled workers. On the other hand, technological development and change have encouraged employers to subdivide a large task into a collection of smaller tasks. This reduced the average level of skill required of the worker. The spread of universal education tended to create a large supply of skilled and semi-skilled workers; whereas technological development, acting via a process of increasing the divisions of labour, tended to lower the level of skill demanded of labour.

According to Reder (1955, p. 217), "mechanisation and specialisation of equipment tends greatly to reduce the need for broadly skilled workers, thereby facilitating the utilisation of partially skilled operatives, whose highly specialised training can be acquired quickly. This tends further to increase inter-skill substitutability." Thus, technological development drives up the wage rates of the unskilled and semi-skilled jobs relative to those in the skilled jobs. A counteracting force was the trend towards higher levels of educational attainment that has increased the number of workers able to compete for the skilled jobs. This puts an upward brake upon the increases in the skilled wage rates. If the aggregate demand for labour increases, there is a tendency for the unskilled to move upwards into better jobs, and thus reduce the supply of labour for the less attractive jobs. This should bid up the wage rate of the less skilled and unskilled. As a result, these counteracting forces have been contributing to the narrowing of the spread between the skilled and the unskilled, or put differently, to the narrowing of occupational, industrial, and regional wage differentials.

One approach that emphasises equalising forces acting on the distribution of personal income was presented in the work of Phelps Brown and Hart (1952). The study gave particular attention to money wage rates and to changes in the number of wage-earners. They showed that over the course of industrial development there was an increase in the number of wage-earners relative to the number of the self-employed, family workers, and small employers. However, the number of wage-earners declined in proportion to earners in other middle level occupations because of the growth in the number of technical and administrative positions. Since these other occupations were considered higher paying, they concluded that this shift from wage-earners to earners in other occupations should tend to raise the median income of the entire population, as well as lower the relative share of wage-earners. In addition, Phelps Brown and Hart examined movements within these other occupations. Assuming a net growth in the labour force, they concluded that because newcomers to these occupations tend to be concentrated in the lower ranks, there should be a slight countering effect pulling down the median income.

Similarly, Denison (1954) added that there has been an overall reduction in the sources of income available to the higher income groups with the result of more positions being available in the middle level occupations. This shift to the middle level occupations, the other occupations of Phelps Brown and Hart, should reduce the overall dispersion of the distribution of personal income. Further considering this kind of analysis, we note that Phillips (1960) also observed this shift to middle level occupations. In this work Phillips pointed out the significance of the movement from self-employed to salaried occupations in reducing the dispersion of the distribution of personal income.

In two other sociological approaches Reder (1955) and Turner (1957) examined specific institutional forces that may be generating the shifts in occupational structure, and hence be causing the narrowing of wage differentials and a less skewed distribution of personal income. On the one hand, Reder (1955) concentrated on the institutional mechanisms that an employer may use to adjust the total level of pay. He suggested that an employer may adjust wage rates, hiring standards, or a combination of the two in discriminating among the applicants on the basis of differences in quality. Recognising that the price of labour (the wage rate) is characteristically sticky or adverse to a fall, and that wage rates are adjusted on a contractural basis at discreet intervals, he concluded that the use of the wage rate adjustments alone is not a sufficiently flexible device for the employer.

Therefore, Reder maintained that adjusting quality or hiring standards offers more flexibility and continuity in downgrading or upgrading workers. In the case of upgrading workers, only the promoted receive an increased pay. However, if the wage rate were to be increased to one without a job reclassification, all workers would have to receive the increase. Equally important is that the adjustment of standards to reclassify workers downwards is less prone to union resistance than an overall wage cut affecting the common cause. Consequently, the adjustment of hiring standards contributes to the narrowing of the differentials.

On the other hand, Turner (1957) attributed the narrowing of the differentials and the decreased dispersion of the distribution of personal income to the rise of mass unionism and its effects on the mechanism determining wages. High employment is accompanied by increasing union membership and by more effective collective bargaining processes, because both the government and the employers are more willing to avoid a disturbance of prosperity. One result of the rise of mass unionism was that the establishment of industry wide agreements which replaced local or occupational agreements.

In this approach, Turner considered that a union can either negotiate wage increases in terms of equal absolute amounts or equal percentage amounts. The smaller craft unions that depended upon apprenticeship to maintain a supply of labour found the spread of mass unionism a threat to their memberships. Hence, they joined the mass unions in the bargaining process to protect apprenticeship and membership. Since large numbers of the less skilled workers tended to dominate these unions, the need to maintain relative differentials subsided. Thus, these unions generally seek flat-rate increases which not only narrowed the differentials, but treated the less-skilled and the lower paid relatively better.

Other unions prefer percentage wage demands which maintain differentials. These unions are usually industry specific and are dominated by the more skilled workers in the industry, such as the unions in the steel industry. They do not rely upon specific apprenticship programs to maintain a supply of labour, and thus, do not seek nor require large numbers of the less skilled in their membership. Turner (1957, p. 239) wrote:

> Such unions have not interested themselves in expanding into other employments: they are, therefore, not compelled by a need to maintain some parity between members in different industries - as are the great craft and general labour unions. But these exceptional unions have no apprenticeship system, and entry to the better paid jobs is by promotion from the less skilled workers. In these cases it seems that the lower paid have been reconciled to the maintenance of abnormal differentials by the chance of ultimately enjoying them.

Even though wage demands may still appear to maintain differentials in cash terms, they are not necessarily maintaining the differentials relatively. Therefore in the Turner analysis, the belief is that the general direction of unionism has been to advance the cause of the less skilled by reducing the relative wage differentials through a preference for flat-rate wage increases. This, then, has given rise to a less dispersed distribution of personal income.

The last approach we shall present is the demand-supply theory or the "tension" theory of the distribution of personal income put forth by Tinbergen (1970). The distribution of personal income is considered to be a result of the supply of and the demand for labour. The demand for labour is characterised by one or more variables describing the type of labour demanded, and the supply is characterised by one or more variables describing the labour supplied. The variables or attributes of the demand for and the supply of labour, such as intelligence, skill level, education, personal characteristics, are called the demand factors and the supply factors. According to Tinbergen (1972,p. 256), "the difference between the values of the demand factors and those of the supply factors, when reduced to some common denominator, can be called the tension between demand and supply quantities."

If the distribution of attributes demanded and supplied were identical, Bjerke (1970, p. 245) noted that "production would be organised in such a way that every job would be performed by the person who was qualified to perform it." However, these distributions are not identical, and tension occurs when the attributes that are supplied are different from those that are demanded. In other words, tension is another way of indicating the relative scarcity of skills supplied relative to the skills demanded. High incomes will be paid when tension is high and lower incomes when tension is low. Therefore, to maximise income an individual supplier will seek to increase his position to the highest tension attainable.

Tinbergen then suggests that income distribution is derived from the distributions of attributes demanded and the attribute supplied. This implies that income distribution depends on tensions. If tensions can be reduced, the inequality of the distribution of personal income can be reduced. Tinbergen (1970, pp. 226-227) adds that "equal incomes are possible not merely if people are equally skilled - which they evidently are not - but already if only the skill distribution required by the organisers of production coincide with the actual distribution."

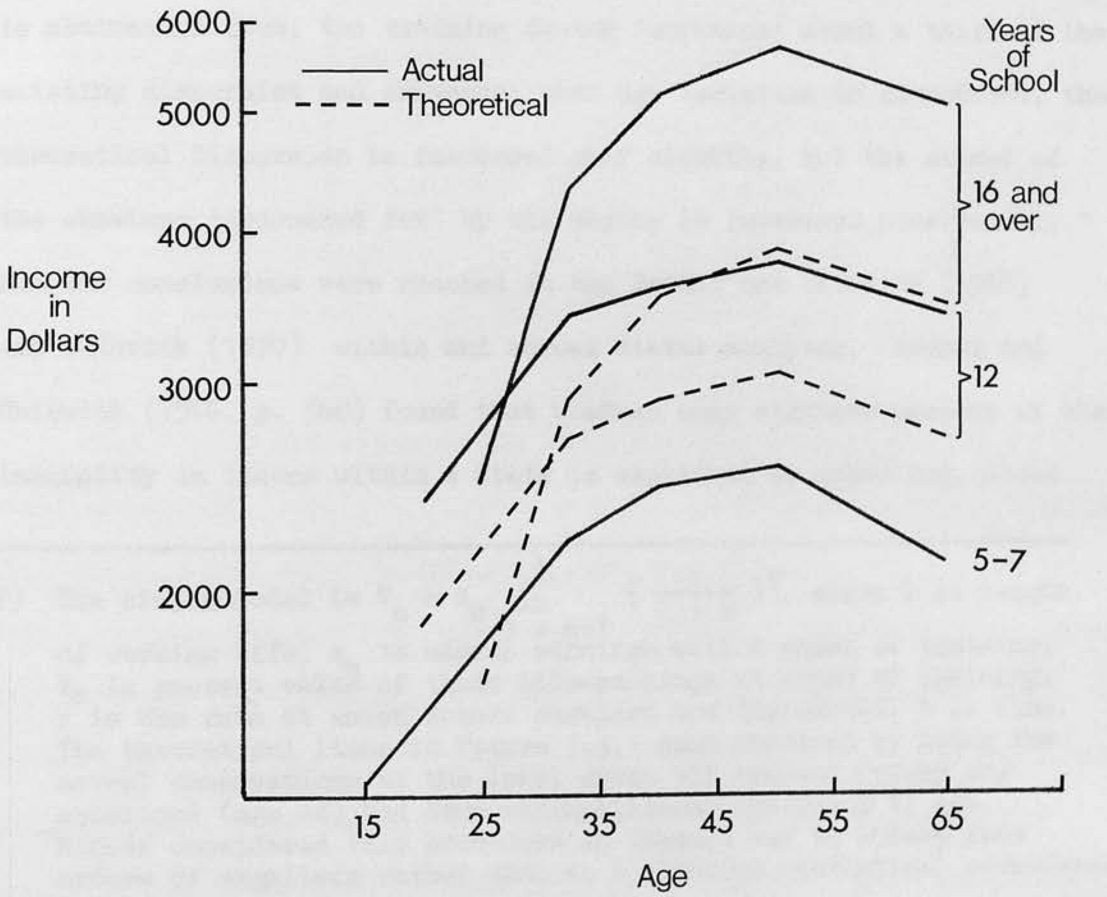
2.4 Related Empirical Work

In this section we only present a selection of empirical works from the vast amount of literature available on empirical analyses of the distribution of personal income. The works included were selected to illustrate the general direction of the empirical studies on the distribution of personal income. This discussion serves to demonstrate the quantitative importance of institutional factors on the distribution and to indicate the general conclusions from empirical research.

Studies by Mincer (1958), Becker and Chiswick (1966), and Chiswick (1970) analysed the human capital approach. This is to say they tested the strength of the human capital concept in explaining the skewness of the distribution of personal income. We present Figure 2.4.1 to illustrate graphically the influence of education, one form of human capital, and age on an individual's earning capacity. Figure 2.4.1 is taken from the Mincer study on US males and shows the actual life path of income and Mincer's theoretical

Figure 2.4.1

Actual and Theoretical Life Paths of Income, U.S. Males 1949.
(Mincer, 1958, p. 294)



life path of income.[2] From the life paths of income in Figure 2.4.1 we can conclude that education is a significant factor in explaining the higher levels of income. In addition, we can infer that incomes are higher with age, but only up to an age of peak income earning capacity.

From an analysis of the coefficients of dispersion and skewness,[3] Mincer (1958, p. 294) concluded "that when age variation of earnings is abstracted from, the training factor 'explains' about a third of the existing dispersion and skewness; when age variation is introduced, the theoretical dispersion is increased only slightly, but the extent of the skewness 'accounted for' by the theory is increased considerably." Similar conclusions were reached in the Becker and Chiswick (1966) and Chiswick (1970) within and across states analyses. Becker and Chiswick (1966, p. 368) found that whereas only eighteen percent of the inequality in income within a state is explained by schooling, about

---

2) The simple model is $V_n = a_n \sum_{t=n+1}^{l} \left(\frac{1}{1+r} \right)^t$, where l is length of working life, a_n is annual earnings with n years of training, V_n is present value of their life-earnings at start of training, r is the rate at which future earnings are discounted, t is time. The theoretical lines in Figure 3.4.1 were obtained by using the actual observations at the level where all present values are equalised (age 14) and from extrapolations according to age. Mincer considered this procedure an inexact way to attain some orders of magnitude rather than as a vigorous statistical procedure.

3) The coefficient of dispersion is the ratio of the difference between the 5th and the 95th percentiles of the distribution to the 50th percentile, $\frac{Q_{95} - Q_5}{Q_{50}}$. The coefficient of skewness is $\frac{(Q_{95} - Q_{50}) - (Q_{50} - Q_5)}{Q_{95} - Q_5}$. The actual measures are calculated from the distribution of incomes of male workers in the United States in 1949; while the theoretical measures are based upon the differences between the appropriate percentiles in the distribution of training (schooling).

one-third of the differences in inequality between states is directly explained by schooling. These studies showed that the influence of schooling alone could produce a positive skewness in the distribution of personal income.

Furthermore, Mincer (1958) analysed three different disaggregations of income distribution. The first was a disaggregation according to age. He partitioned the income recipients into age groups and demonstrated that there is an increased dispersion of income with an increase in age. The second disaggregation grouped the income recipients by educational background, and demonstrated that income dispersion increases with the level of education. The last disaggregation he presented grouped the income recipients by occupational class. The higher the level of skill demanded in the occupation, the greater was the dispersion in the income distribution.

Keat (1960) investigated the trend in percentage wage differentials among occupations requiring different skill levels. His findings confirmed that the relative skill differentials had fallen during the period 1900-1956. He then proceeded to investigate separately and explicitly a few institutional and socio-economic factors that may have influenced the occupational wage structure. The factors he analysed were sex and racial differentials, union membership, immigration, age, and education and training. From his analysis Keat attributed a significant part of the explanation of the narrowing of wage differentials to education and training, but only minor importance to the other factors in explaining the occupational wage structure.

In a study of the narrowing of income differentials Soltow (1960) analysed the influence of three long run trends: firstly, the dispersion of educational attainment; secondly, the occupational distribution; and thirdly, the age distribution. The procedure adopted was to compute three coefficients of concentration by relating changes in the dispersion of income firstly with shifts in the dispersion of education, secondly with the shifts in the dispersion of employment among the occupations, and thirdly with shifts in the dispersion of age.[4] Soltow presented the three coefficients in index form as shown in Table 2.4.1.

From the evidence in Table 2.4.1 Soltow concluded that age shifts have had some effect in increasing inequality. On the other hand, shifts in occupations towards those with less income dispersion and upwards shifts in the educational levels were strong factors in decreasing the inequality of income. From Table 2.4.1 we can observe shifts in the importance of the occupational distribution were historically more significant in reducing the inequality of the distribution of personal income than shifts in educational attainment.

Regional and state studies by Adams (1958), Al-Samarrie and Miller (1967), and Farbman (1973) sought to explain the distribution and the concentration of personal income by models formulated in terms of both economic and non-economic variables. Each study used regression

---

4) The coefficient of concentration is the Gini coefficient measured here as the mean difference in income divided by two times the arithmetic mean, that is,

$$R = \frac{1}{2\,\bar{X}} \left[\frac{\sum_{r,s=1}^{m} |X_r - X_s|\, f_r\, f_s}{N^2} \right]$$ where the X's are the income variates and the f's refer to classes, depending upon the characteristic.

Table 2.4.1

Income Coefficients of Concentration Indexes Obtained by Relating the Changes in the Distribution of Age, Major Occupation, and Education to the 1956 Distribution of Income (1956 = 100) (Soltow, 1960, p. 453)

Year	Age	Occupation	Education
1900	94	123	
1910	94	118	
1920	95	114	
1930	96	110	
1940	98	109	104
1950	99	103	103
1960	101		100
1970	102		97
1980	101		95

analysis and analysis of variance. Income was considered dependent on economic and non-economic variables. Adams (1958, p. 395) analysed income as a function of six variables - education, occupational structure, age, location, community size, and part of the year worked. Al-Samarrie and Miller (1967, p. 63) employed five variables - proportion of property income in personal income, agricultural weight in the economy, education, racial discrimination, and public sector employment. Farbman (1973, pp. 335-36) used six variables - racial discrimination, education, activity rates, occupational structure, mean family income, and urban/rural make-up.[5] Each of the three studies succeeded in explaining about half of the inter-state variations in the distribution of personal income by economic and non-economic variables. In the three studies the more significant variables contributing to an explanation of the distribution were racial discrimination, occupational structure, urban/rural make up, and education. Adams (1958, p. 396) noted that for the higher income groups the residual is positively skewed, and for lower income groups the residual is normally distributed. These studies concluded that the residual may be dependent on income in the previous period, which is a condition leading to the lognormal distribution. These kinds of studies on small geographic areas serve to identify the factors within each area which significantly affect incomes and to identify the variations in the significance of factors from area to area.

---

5) The measure used for racial discrimination in both the Al-Samarrie and Miller and the Farbman studies is the per cent of the resident population which is non-white.

The last study to be discussed in this section is similar to the Adams, Al-Samarrie and Miller, and the Farbman studies in that it relies upon a simple model of individual earnings. This study we refer to is that of Hanushek (1973) which analyses individual male earnings as a function of education, ability, and experience. Hanushek used a survey of enlisted men leaving the U.S. Army in 1969 because this provided more detailed information on training, ability, and experience than was available in the United States census. A central part of his analysis was to examine the differences in earnings by regions.

From the analysis Hanushek (1973, p. 210) observed that "the differences in returns to human capital are 'pure' regional effects rather than further manifestations of the embodied human capital in each individual." He found that almost sixteen percent of the variance in earnings resulted from differences in the mean earnings among the regions. He tested whether the differentials in mean earnings were a reflection of the input differentials or the structure of earnings. He concluded that structural differences accounted for more than eighty percent of the variance in mean earnings. From this, Hanushek (1973, p. 210) recommended "that more effort should be devoted to analysing the structure of the labour market than looking at the distributions of individuals and their characteristics in analysing income patterns."

2.5 <u>Summary</u>

The theoretical-statistical approach is significant for describing the stability or regularity of the distribution of personal income and its positive skewness. The approach expresses the distribution of personal income in terms of statistical distribution laws. However, in fitting statistical frequency functions to the distribution of personal income, the theoretical-statistical approach only <u>describes</u> the characteristics of the distribution. It maintains that anonymous random forces generate the distribution. The approach may offer simplicity and mathematical elegance, but not positive and firm explanations of the underlying causes of the distribution of personal income.

The sociological approach attempts to explain the distribution of personal income through analysis of many institutional forces. It concentrates upon the individual components which may influence the distribution. The shape of the distribution curve is thought to be influenced by the strength of the various factors. The empirical works summarised above demonstrated the quantitative significance of isolating some of the economic and institutional factors.

Nevertheless, the identification of several significant factors at work in the distribution process does not in itself guarantee a complete and adequate explanation, Bjerke (1970, p. 247) wrote:

> Some of the sociological views of the importance of the institutional factors nevertheless do not give any proper explanation because these institutional factors are themselves a result of certain processes. It may therefore be asked what are the underlying forces which have produced the sociological factors which influence income distributions. A dynamisation of the conceptual apparatus of the sociological school seems to be lacking.

In the final analysis, the distribution process cannot be isolated from the income creation process. Neither the theoretical-statistical nor the sociological approaches concentrated on the underlying manner in which the income was acquired. The mode in which incomes accrue was left as a secondary factor, and therefore, the mechanism which contributed to the formation of the distribution of income was not described. The approaches discussed in this chapter did not consider the actions or the mechanism of the market structure facing the individual. The approaches are independent of the limitations on individual choice imposed by the market structure. However, the choices are determined in a market whose structure and function contribute to the acquisition of personal income and the formation of the distribution of personal income. Constrained as they are, individual choices, nevertheless, can only be appreciated in a market of known structure and operation. It is the purpose of the subsequent chapter to investigate the earnings process within the context of such markets.

CHAPTER 3

The Labour Market

3.1 The Labour Market and the Distribution of Personal Income

The literature on the distribution of personal income shares presented in Chapter 2 depicts the distribution process in terms of either general stochastic processes or sociological and institutional forces influencing the income level of an individual. Neither approach to the distribution of personal income treated the specific activities and conditions under which the labour market influences the allocation and the pricing of labour. The conditions and the operations of the labour market determine the position of an individual in the labour market. This position in the labour market reflects the motives and the scope for action of the individuals, as well as the behaviour of enterprise, in the process of creating and distributing personal income shares.

In other words, we must consider the relation between the creation and the distribution of personal income. We regard the labour market as a link between income creation and income distribution: in a sense a link between the microeconomic activities of production and the macroeconomic outcome of distribution. Nevertheless, this link between the creation and distribution of personal income is not straightforward. For example, as we saw in the human capital approach in Chapter 2, even though two individuals may possess the same potential to obtain equal economic rewards, one may ultimately earn more than the other. The human capital approach attributed this outcome to an individual's "economic talent" in exploiting his human capital.

However, through barriers to mobility or limitations on the economic structure, the operations and structure of the labour market might inhibit one of the individual's efforts to exploit his human capital, whilst encourage the other.

The individual return to labour is determined through the operations of the labour market. To ignore the operations of the labour market is to ignore the organisation of the factors of production and the pricing of the factors. The labour market is influenced by a mixture of social, institutional, and economic factors. There are employer associations, employee associations, and government institutions - each of which acts to formulate rules and precedents and to exert influence on the operations of the labour market.

The labour market is like other markets in which buyers and sellers are organised to determine the terms on which transactions proceed. In the labour market the commodity being offered for trade is the services of men and women. The sellers in the labour market are active agents with opinions about the operations of the labour market. Therefore, the social and institutional environment in which the labour market is organised is as important as the economic environment in determining the operations of the labour market. The non-market forces bearing upon the buyers and sellers in the labour market should be included in a realistic presentation of the labour market.

The primary function of the labour market is the allocation of labour within the market. In this thesis we are concerned with the way different kinds of labour are allocated to the jobs in the labour market. In a commodity market, the means of allocation is the price mechanism by which changes in demand are reflected in price changes. However, in the labour market, allocation and income depend upon the individual's decision to supply labour, the structure of the labour force, and the system of prices determined by social and institutional forces. Lydall (1968, pp. 3-4) pointed out the relation between personal income and the market structure: "The structure of earnings in a competitive market can be expected to reflect both the conditions of supply of labour services of different qualities and the conditions of demand, which, in turn, depend on the level of technique, the supply of other factors and the pattern of market demand for the products."

Consequently, the purpose of this chapter is to present a survey of the explanations of the earnings process in the labour market. In this chapter an emphasis is placed upon the importance of determining the factors in the labour market that not only explain the earnings process but also an individual's position in this process. Therefore, we present the discussion of the labour market to establish the importance of a concept of the labour market that reflects the influences of social, and institutional factors, as well as economic factors. Firstly, the discussion will present the conventional approach to the labour market. The conventional approach relies on the idea of marginal productivity and general equilibrium analysis to explain the structure and operations of the labour market in the determination of wages. Secondly, the discussion will consider the adaptability of the conventional approach in accommodating revisions required for the recognition of elements of imperfect competition in the market. Thirdly, the discussion will discuss the development of alternative models of the labour market that reflect the influences of other real world factors in the labour market.

3.2 The Conventional Approach to the Labour Market

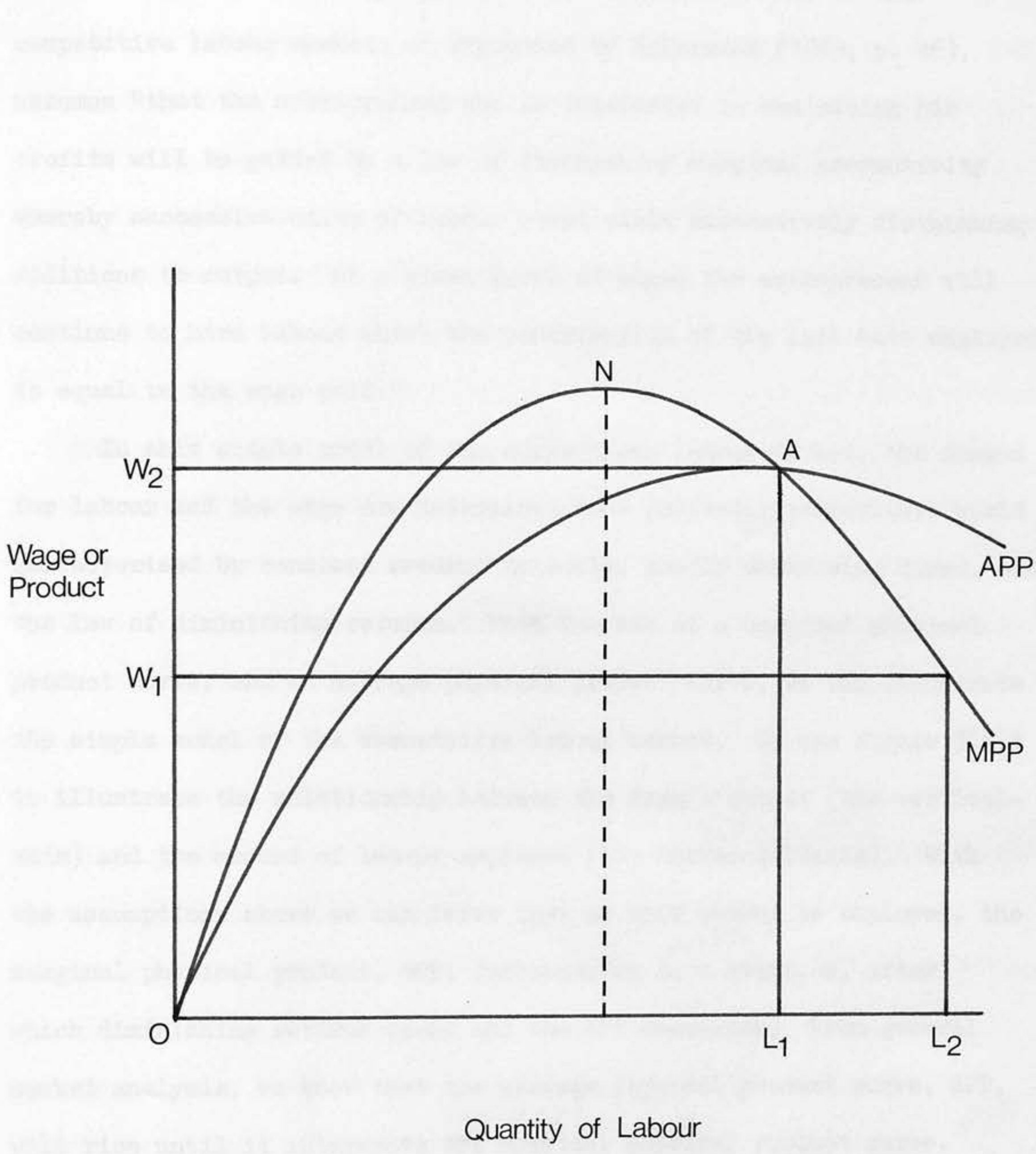
The conventional concept of the labour market is an extension of general market theory, that is, the labour market is a market mechanism allocating labour and determining labour's share according to marginal productivity. This view of the labour market is based upon the concept that market forces act through the price (wage) mechanism to adjust the demand for and the supply of labour. We shall refer to this labour market as the competitive labour market. There are some necessary assumptions adapted from general market theory that must be stated before discussing the competitive labour market. Firstly, there is a large number of employers (buyers) and employees (sellers) in the market. Secondly, labour is homogenous. Thirdly, employees do not prefer one employer to another. Fourthly, neither employers nor employees have influence over the wages. And fifthly, both employers and labour have perfect knowledge of job vacancies, opportunities, and wages.

The competitive labour market recognises different types of labour, characterised by differences in earnings, but not classes of labour, and subjects all types of labour to the same conditions. The structure of the competitive labour market recognises rational, consistent, and economic behaviour on the part of employers and labour. The physical relations governing supply and demand regulate the price (wage) of labour so that no excess of demand or supply will exist. The labour market is cleared of all types of labour by these same physical relations.

In the simple model of the competitive labour market, the demand for labour is a derived demand, and the supply of labour is considered perfectly elastic. The demand for labour is derived from

Figure 3.2.1

Competitive Model of Labour Market



the contribution labour makes to production and the demand for the products which labour helps to produce. The demand side of the competitive labour market, as expounded by McCormick (1969, p. 46), assumes "that the entrepreneur who is interested in maximising his profits will be guided by a law of diminishing marginal productivity whereby successive units of labour hired yield successively diminishing additions to output. At a given level of wages the entrepreneur will continue to hire labour until the contribution of the last unit employed is equal to the wage paid."

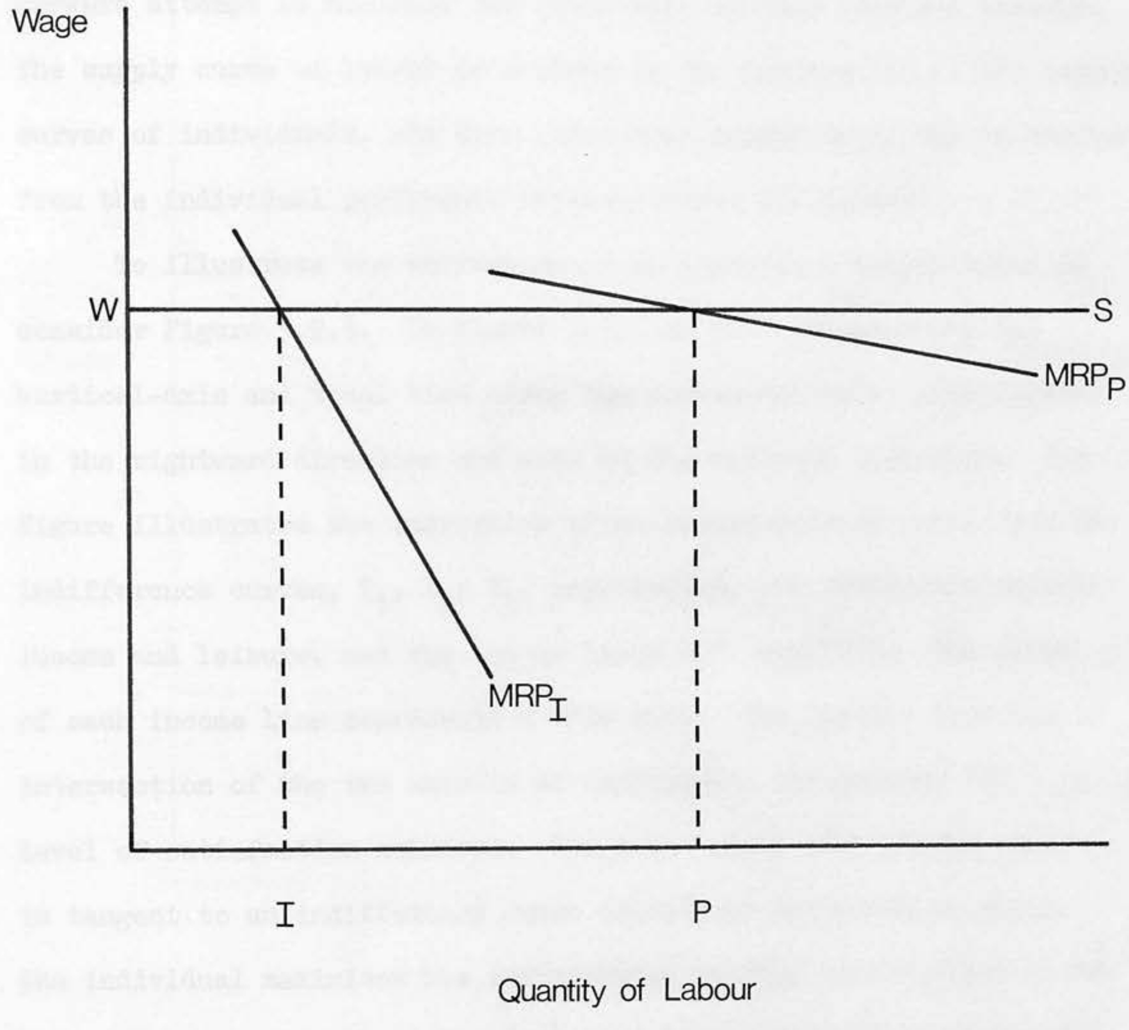
In this simple model of the competitive labour market, the demand for labour and the wage are determined in a perfectly competitive world characterised by constant returns to scale, profit maximising firms, and the law of diminishing returns. With the aid of a marginal physical product curve, and an average physical product curve, we can illustrate the simple model of the competitive labour market. We use Figure 3.2.1 to illustrate the relationship between the firm's output (the vertical-axis) and the amount of labour employed (the horizontal-axis). With the assumptions above we can infer that as more labour is employed, the marginal physical product, MPP, increases up to a point, N, after which diminishing returns occur and the MPP decreases. From general market analysis, we know that the average physical product curve, APP, will rise until it intersects the marginal physical product curve, point A in our example. It is at this point that average product per additional unit of labour decreases.

Once the MPP curve for labour is known, we can establish the demand curve for labour. From the conditions of profit-maximisation under perfect competition, we know that the wage can be set up to the point at which MPP $\leq$ APP in equilibrium. Labour is engaged up to the point A at which the wage of the last worker equals the output produced by the last worker. If the wage is set at W_2, the employer will employ the quantity of labour L_1, and thus maximise profits. Even if the wage falls to W_1, the employer can still earn profits by hiring more labour, and thus L_2, will be employed. Only as the wage falls below MPP = APP is the employer receiving profits, hence the demand curve for labour is the MPP curve below point A. For any rise or fall in the wage, the quantity of labour employed depends upon the slope of the MPP curve.

We should note that differences in the type of competition in the product market will influence the demand for labour. To demonstrate this, we shall use Figure 3.2.2. Firstly we assume that a firm moves from a perfectly competitive product market to an imperfectly competitive product market. In both product markets the marginal physical product curve for labour is the same. Let the curve WS denote the supply curve of labour. Under conditions of perfect competition, any additional output can be sold at a constant price. Therefore, since the marginal physical product curve is downward sloping, we can represent the demand for labour by the marginal revenue product curve, MRP_p in Figure 3.2.2, as downward sloping. Under perfect competition the level of employment in equilibrium is P at wage W. In imperfect competition the additional output requires that the price over the whole output is lowered. The demand for labour is represented by MRP_I. However, in this case, with the same supply curve for labour, we can conclude from Figure 3.2.2 that the level of employment in equilibrium is I, which is less than P from the perfectly competitive situation. Put differently, depending upon the type of competition in the product market, the elasticity of demand for the product will generate rises or falls in the demand for labour. Thus, the quantity of labour employed is determined in part

Figure 3.2.2

Demand for Labour Under Perfect and Imperfect Product Markets



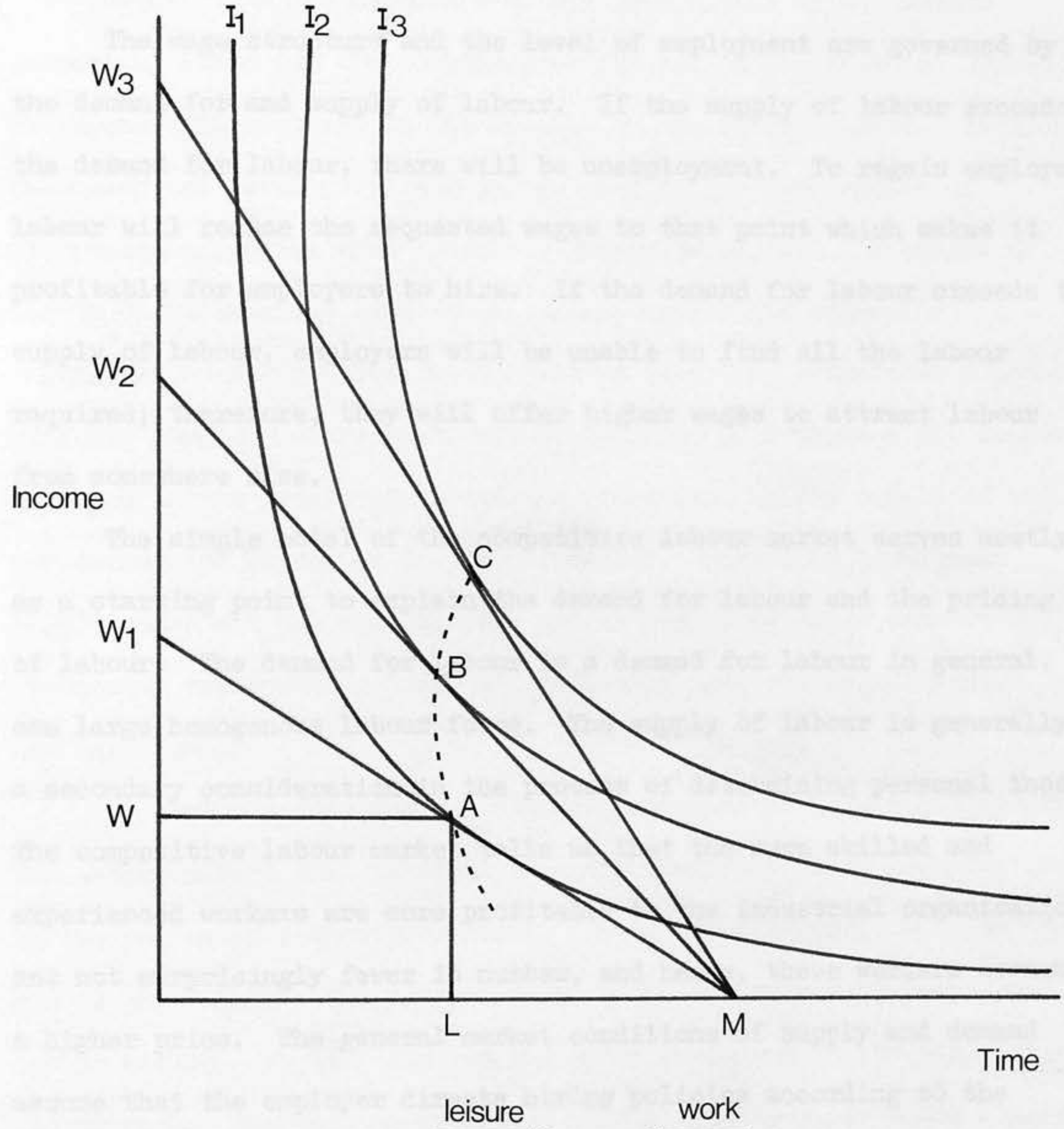
both by the elasticity of demand for labour and the elasticity of demand for the product.

In practice the supply curve of labour is not perfectly elastic, and it may be useful to adapt the competitive model to accommodate this. The notion of worker's choice between work (income) and leisure determines the supply curve of individual workers. In other words, workers attempt to maximise the trade-offs between work and leisure. The supply curve of labour as a whole is the aggregation of the supply curves of individuals, and each individual supply curve may be derived from the individual preference between income and leisure.

To illustrate the derivation of an individual supply curve we consider Figure 3.2.3. In Figure 3.2.3 we plot income along the vertical-axis and total time along the horizontal-axis, with leisure in the rightward direction and work in the leftward direction. The figure illustrates the derivation of an income-leisure curve from the indifference curves, I_1, I_2, I_3, representing the trade-offs between income and leisure, and the income lines W_1M, W_2M, W_3M. The slope of each income line represents a wage rate. The further from the intersection of the two axis is an individual, the greater the level of satisfaction achieved. The point where each income line is tangent to an indifference curve represents the point at which the individual maximises his satisfaction on that indifference curve. For example, on I_1, the individual maximises satisfaction by having OL leisure time and OW income, or put differently, taking OL leisure time and LM working time. Joining the points of maximum satisfaction on the three indifference curves, A,B,C, produces the income-leisure curve. Labour will be supplied up to the point at which the marginal rate of substitution of income for leisure is equal to the wage rate.

Figure 3.2.3

Supply of Labour



In Figure 3.2.3 the labour supplied increases with income, but after point B more leisure is preferred. In other words, after B the utility of the wage is no longer sufficient to encourage the substitution of income for leisure.

The wage structure and the level of employment are governed by the demand for and supply of labour. If the supply of labour exceeds the demand for labour, there will be unemployment. To regain employment labour will reduce the requested wages to that point which makes it profitable for employers to hire. If the demand for labour exceeds the supply of labour, employers will be unable to find all the labour required; therefore, they will offer higher wages to attract labour from somewhere else.

The simple model of the competitive labour market serves mostly as a starting point to explain the demand for labour and the pricing of labour. The demand for labour is a demand for labour in general, one large homogenous labour force. The supply of labour is generally a secondary consideration in the process of determining personal income. The competitive labour market tells us that the more skilled and experienced workers are more profitable to the industrial organisation, and not surprisingly fewer in number, and hence, these workers command a higher price. The general market conditions of supply and demand assume that the employer directs hiring policies according to the potential productivity of a worker. In the competitive labour market analysis it is assumed that the employer is a profit maximiser and uses productivity to evaluate the potential employees upon strictly economic terms. The system envisaged operates freely, competitively, and completely. Thus, we find in the simple model of the competitive labour market that the pricing of labour eliminates any excess demand or supply.

3.3 Imperfections in the Conventional Labour Market

However, as a consequence of the development of the modern industrial state and the accompanying changes in the economic system, revisions were required in the concept of the labour market. Advanced technological development and change lead to the development of production processes that assembled groups of sellers and buyers of labour in centralised locations. Transporation and communication encouraged centralisation, made movement easy, and stimulated a common outlook. Employers associations, unions, and government - all became involved in the interactions and the operations of the supply of and the demand for labour in the labour market. According to Marchal (1957) wages and the wage process became perceived as an element of production representing power and influence over the labour market and the economy.

We must now consider the changes or revisions in the concept of the labour market that resulted from the introduction of imperfections in the perfectly competitive model. These imperfections, both on the demand and supply sides, interfere with the operations of the competitive labour market. We will discuss the major revisions to conventional labour market analysis as set forth in Dunlop (1950), Ferguson (1972), Hunter and Robertson (1969), Jackson (1970) and others. To begin, we examine the situation in which a few profit-maximising employers control a large portion of the jobs in the labour market. This is a demand side imperfection. We shall treat these employers as though they were one. In other words, this is a situation in which there is monopsony in the labour market. In the first instance we also assume the employer is a monopolist in the product market. We assume that any change in the demand for labour by the monopsonist will affect the wage rate and the market demand for labour.

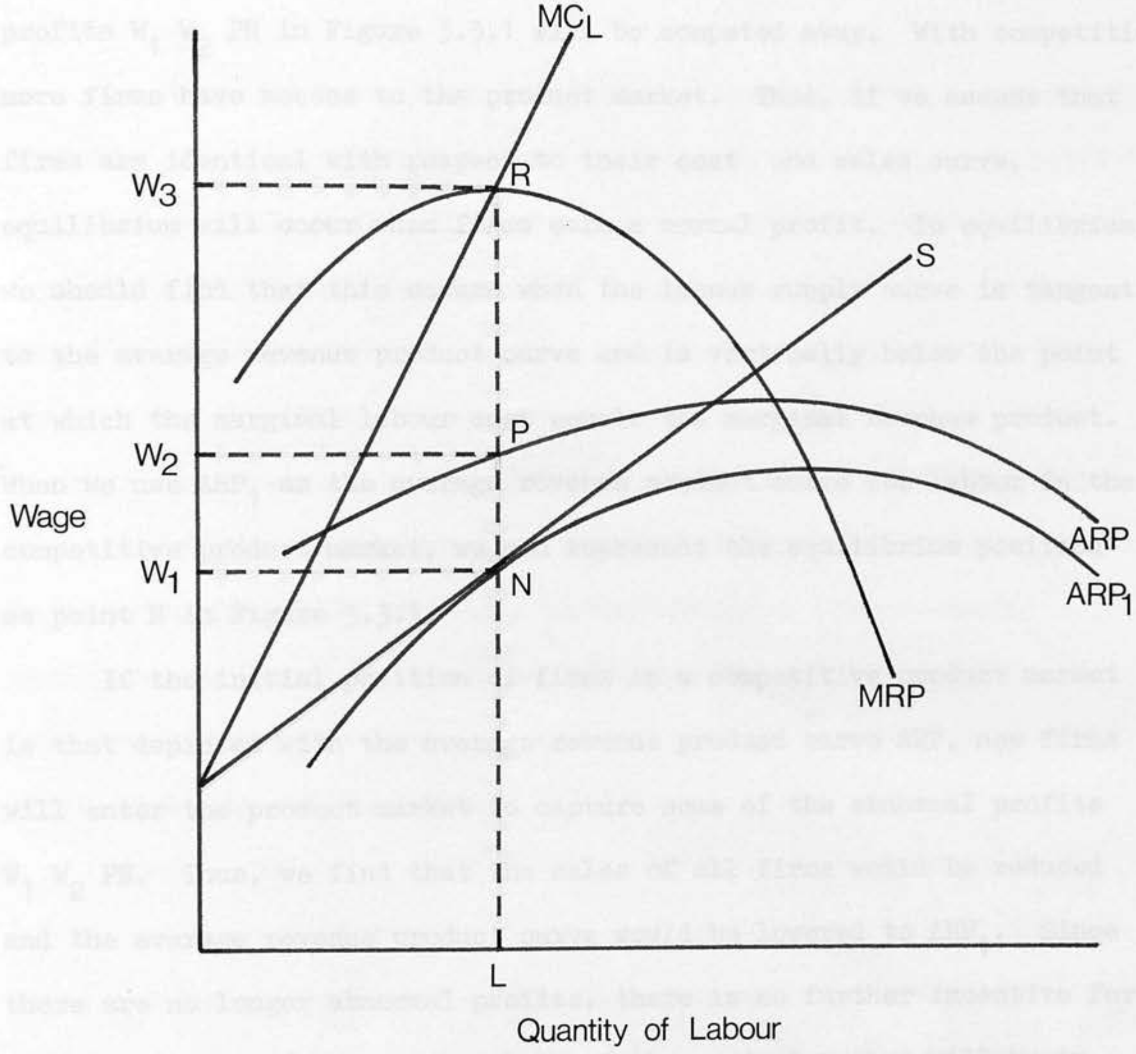
Under the conditions of monopsony, any increase in labour demand must be accompanied by an increase in wages to motivate the required supply. Figure 3.3.1 provides an illustration of the ffects of monopsony in the labour market. In Figure 3.3.1, S represents the supply curve of labour. The curve shows that to attract more labour, a higher wage must be offered. We let the average revenue product curve by ARP and the marginal revenue product curve for labour by MRP in Figure 3.3.1. Since the same wage is generally paid to all workers, the monoponist must raise the wages for all the labour employed. The curve MC_L represents the marginal labour cost or the amount by which the wage bill rises with the addition of one worker.

In equilibrium, the monopsonist will employ additional workers up to the point at which the marginal revenue product for labour is equal to the marginal labour cost, point R in Figure 3.1.1. If any more labour is employed, the marginal revenue product is less than the cost, and hence, the monopsonist is losing money. From Figure 3.3.1, we can conclude that to employ L workers, the employer must offer a wage of LN. The marginal cost of employing the L-th worker is shown by LR. In this example using the curve ARP as the average revenue product of labour, we find that when L workers are employed the average revenue product of L workers is PL. This is greater than the wage LN, and therefore, the monopsonist realises abnormal profit, W_1 W_2 PN.

Even though an employer may be a monopsonist in the labour market, he is not necessarily a monopolist in the product market.

Figure 3.3.1

Monopsony in the Labour Market



In the example discussed above, we used the curve ARP as the average revenue product curve. This is an example of an employer that is a monopsonist in the labour market and a monopolist in the product market. However, if we consider a monopsonist in the labour market that faces competition in the product market, we should find the abnormal profits W_1 W_2 PN in Figure 3.3.1 will be competed away. With competition more firms have access to the product market. Thus, if we assume that firms are identical with respect to their cost and sales curve, equilibrium will occur when firms earn a normal profit. In equilibrium, we should find that this occurs when the labour supply curve is tangent to the average revenue product curve and is vertically below the point at which the marginal labour cost equals the marginal revenue product. When we use ARP_1 as the average revenue product curve for labour in the competitive product market, we can represent the equilibrium position as point N in Figure 3.3.1.

If the initial position of firms in a competitive product market is that depicted with the average revenue product curve ARP, new firms will enter the product market to capture some of the abnormal profits W_1 W_2 PN. Thus, we find that the sales of all firms would be reduced and the average revenue product curve would be lowered to ARP_1. Since there are no longer abnormal profits, there is no further incentive for more new firms to enter the market and the product market will be in equilibrium.

A second type of imperfection resulted from the rise of the trade and industrial unions. Unions could interfere with the competitive operations of the labour market by imposing restrictions on the supply of labour. Through unions, individuals could act collectively in the labour market. Unions served to increase bargaining power. By directly withdrawing labour in industrial disputes or indirectly limiting membership, unions could exert influence in the wage determination process. Since the supply of labour, to a considerable extent, was represented by unions instead of individual wage earners, the concept of the labour market was revised to accommodate the economic effects

If we assume that unions monopolise a large portion of the supply of labour, we can no longer consider the supply of labour to be perfectly elastic. We still assume that the employers are profit-maximisers and that their behaviour is as described in the discussion above on the competitive labour market. In this situation the union can then use one of three possible alternatives to raise the wage. One is to restrict the supply of labour either by directly limiting membership or by indirectly raising job requirements. A second alternative is to restrict the supply of labour by negotiating a wage at which employers are forced to reduce employment. The third alternative is to increase the productivity of labour which will generate an upward shift in the demand for labour.

However, if we assume that unions do monopolise the supply of labour, we must realistically assume that in following one of the three alternative courses of action the union seeks to achieve or maximise some objective. The difficulty is stating this objective. On the one hand, Dunlop (1950, p. 36) considered that the overall objective is "to achieve the largest possible wage bill from the particular segment of the economy, quite regardless of whether all wage earners are employed." On the other hand, Ross (1948) regarded the objective as not necessarily the maximisation of wages or membership or any other factor, but institutional survival and growth. Even though it is difficult to establish precisely the overall objective of a union, in the context of this chapter it is sufficient to consider the basic economic effect of a union. In effect, any of the alternative courses of action available to unions results in the maintenance of a minimum wage. Thus, we shall limit this discussion of monopoly in the labour market to the economic effects of unions maintaining a minimum wage regardless of the overall objectives of unions.

We assume that there is a large number of unorganised employers in the labour market and that these employers compete in a competitive product market. The introduction of the monopolistic seller of labour in the labour market does not alter the simple model of the competitive labour, as illustrated in Figure 3.2.1. No matter how the wage is determined, by the market or the union, employers will only employ labour at a wage which corresponds to a point at which the marginal revenue product is less than or equal to the average revenue product. If we refer back to Figure 3.2.1, the maximum wage is W_2, which is determined by point A at which MRP = ARP. The level of employment is L_1. If labour agrees to the wage W_2 as a minimum, the supply of labour would be represented by the curve W_2A. Employers cannot hire labour below the wage W_2 and cannot profitably employ labour at a wage set above W_2 without reducing the level of employment.

Of course, it is unrealistic to assume that the employers are unorganised. Therefore we must further consider a labour market in which both the buyers and sellers of labour are organised. This is the case of bilateral monopoly. The typical case of bilateral monopoly in the labour market occurs when the supply of labour is dominated by unions and the demand for labour is determined by a single employer or an association of employers.

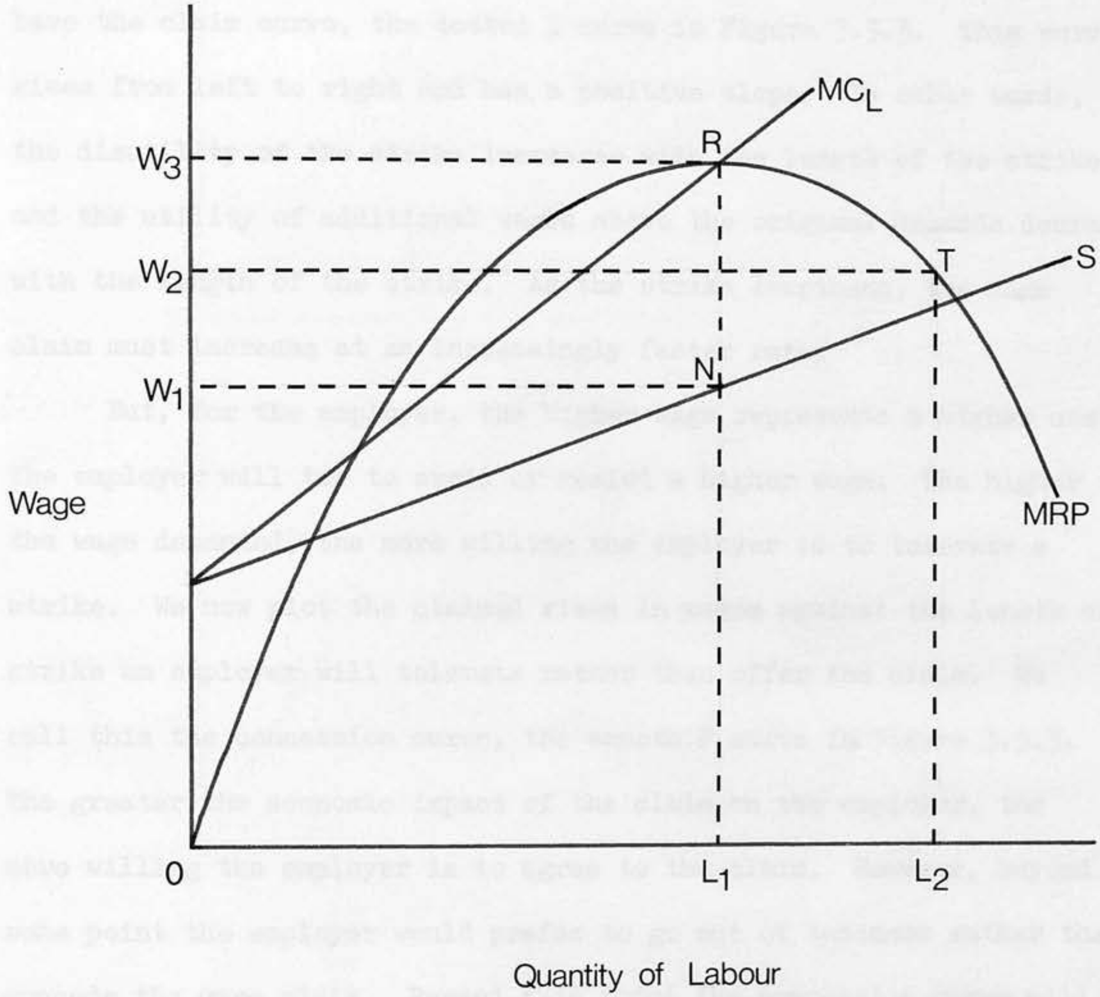
To analyse bilateral monopoly in the labour market we refer to Figure 3.3.2. We let the demand for labour be represented by the marginal revenue product curve MRP, and the supply curve of labour be S. Since the monopsonist is the only buyer, he would normally offer employment up to the point at which the marginal revenue product of labour is equal to the marginal labour cost, point R in Figure 3.3.2. The monopsonist offers the wage W_1 to L_1 men. However, since the union is the only seller, it can make any realistic claim within the boundaries of two courses of action. The first is to comply with the monopsonist's demand for L_1 men, but only at a minimum wage of W_3, which is above the wage W_1 offered. The second is to seek a higher level of employment determined by the intersection of supply and demand. The intersection is T, and the corresponding level of employment is L_2. At this intersection both the wage W_2 and level of employment are more than that offered by the monopsonist.

The general conclusion from this analysis of bilateral monopoly is that the outcome is indeterminate. The scope of the monopsonist's ability to offer any wage above W_1 depends upon the type of product market as was discussed in the explanation of monopsony illustrated in Figure 3.3.1. The solution relies upon either or both of the participants in the labour market changing their courses of action.

There have been some attempts to reach a determinate outcome of bilateral monopoly. A number of these attempts have sought a solution through models of collective bargaining in the labour market. Generally, these approaches themselves are indeterminate. For example, we can use an approach similar to that of Turner (1933). In this approach to bilateral monopoly we assume that unions press demands for higher wages, and employers seek to keep these demands as low as possible.

Figure 3.3.2

Monopoly in the Labour Market



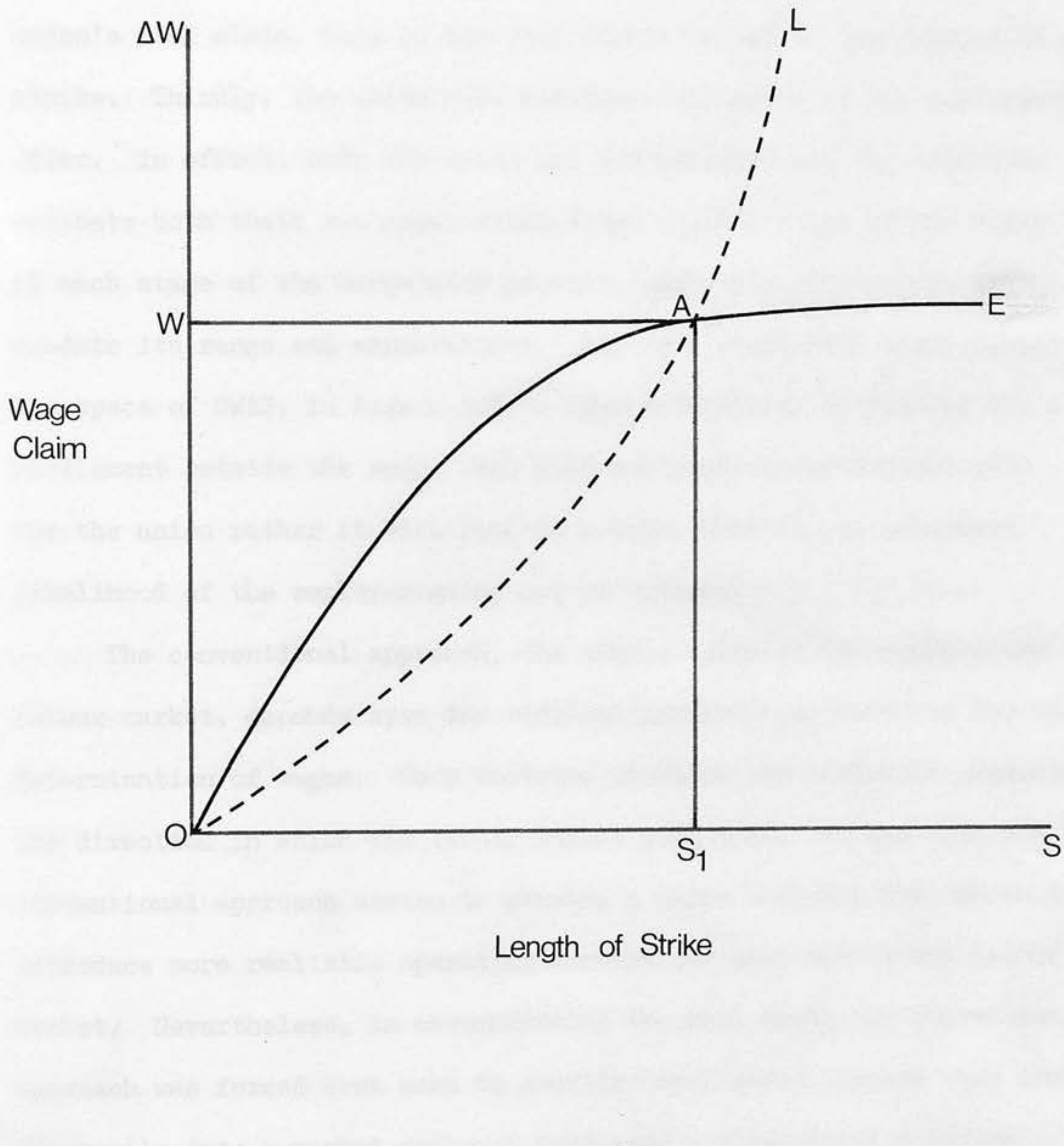
We assume that unions will be willing to go out on strike to secure higher wages. This will incur hardship and financial loss on members. The length of a strike will depend upon the size of the wage claim. The higher the wage claim, the more willing the workers will be to prolong the strike. If we plot the claimed rise in wages against the length of strike worth enduring to secure the rise, we have the claim curve, the dotted L curve in Figure 3.3.3. This curve rises from left to right and has a positive slope. In other words, the disutility of the strike increases with the length of the strike, and the utility of additional wages above the original demands decreases with the length of the strike. As the strike lengthens, the wage claim must increase at an increasingly faster rate.

But, for the employer, the higher wage represents a higher cost. The employer will try to avoid or resist a higher wage. The higher the wage demanded, the more willing the employer is to tolerate a strike. We now plot the claimed rises in wages against the length of strike an employer will tolerate rather than offer the claim. We call this the concession curve, the smooth F curve in Figure 3.3.3. The greater the economic impact of the claim on the employer, the more willing the employer is to agree to the claim. However, beyond some point the employer would prefer to go out of business rather than concede the wage claim. Beyond this point the concession curve will flatten out.

The solution is achieved at the point at which the claim and concession curves intersect, A in Figure 3.3.3. This in itself does not provide any concrete solution as we shall see in the ensuing discussion. The next step in this type of approach as set forth by Johnston (1972b) is to estimate the claim and concession curves as

Figure 3.3.3

Bilateral Monopoly: Duration of Strike and Wage Claim



functions of the costs of the strike and the benefits of the claim. The functions depend on financial and psychological factors. The two bargaining units embark upon a process in which each attempts to balance the gains and losses involved in reaching a settlement. Firstly, the union will establish a claim. Secondly, the employer will estimate cost and benefit functions of the union to arrive at an estimate of the union's real claim, that is the wage needed to reduce the chances of a strike. Thirdly, the union will establish estimates of the employer's offer. In effect, both the union and the employer use the functions to estimate both their own negotiating range and the range of the other. At each stage of the bargaining process, each side will revise and up-date its range and expectations. Any settlement will occur within the space of OWAS, in Figure 3.3.3. Any possibility of pushing for a settlement outside the space OWAS will not lead to any further gain for the union rather it will lead to a loss, that is, an increased likelihood of the employer going out of business.

The conventional approach, the simple model of the competitive labour market, depends upon the marginal productivity doctrine for the determination of wages. This doctrine provides the tools for predicting the direction in which the labour market will tend. We saw that the conventional approach serves to provide a basic analysis from which to introduce more realistic operating assumptions that affect the labour market. Nevertheless, in accommodating the real world the conventional approach was forced even more to consider real world factors that could not fit neatly into a market analysis producing a determinate solution. Therefore, we have reached the point at which it is necessary to deviate further from the limiting assumptions of the conventional approach. In doing so, we will introduce more real world factors that may influence the structure and operations of the labour market.

3.4 Alternative Approaches to the Labour Market

The real world is more complicated than the world described by conventional labour market analysis. It is uncertain with divergent possibilities for education, training, skill development, and geographic mobility. The specialisation of labour and the modern industrial organisation introduced distinctions into the production process. These distinctions resulted in more efficient and economical processes requiring different levels of education and skill. The production processes demand numerous divisions of labour. This splitting of the production process into minutely distinct tasks fragments the labour force. As a result, occupational specialisation is a characteristic feature of individuals in the labour market, and it is this that creates demands for higher levels of investment in education and training which in turn raises expectations for higher returns and specialisation. For these reasons we turn to trace the development of concepts of the labour market that introduced social and institutional determinants into the analysis. These concepts embraced the structural relations in an institutional framework determined by social criteria and power relations among groups. Corina (1972, p. 5) calls this type of labour market an 'atypical' labour market.[1] We shall consider an 'atypical' labour market as alternative to the conventional concept in general market theory. We regard the labour market as an

---

1) The nomenclature used by Corina serves to identify a labour market analysis that differs from the conventional labour market in economic analysis. He does not mean that his nomenclature only applies to an analysis of atypical markets in practice. Indeed, he presumes that 'atypical' labour market analysis applies to an analysis of the typical labour market in practice.

institutional framework reflecting economic, as well as non-economic behaviour. It is an institutional framework embracing attitudes, value systems, and power relationships among groups within the labour market. The emphasis in the 'atypical' labour market is on the actions and scope for actions of the supply of labour. Work rules, formal and informal relationships at the work place, trade union activity, and social inertia are inherent and recognised as part of the labour market structure.

This alternative type of labour market is an outcome of a socio-economic approach that bases the analysis of the labour market on the identification of dominant social and economic groups and the conditions under which the groups act. This labour market is characterised by three features that determine the allocation and pricing of labour: firstly, groups of buyers and sellers; secondly, the conditions of entry and exit in the market; and thirdly, similarities between individual workers and differences in the treatment of similar individuals. In addition, we find that under 'atypical' labour market analysis one no longer assumes that labour is homogeneous. Many broad subdivisions of the labour force, such as the skilled, the semi-skilled, and the unskilled are recognised. These subdivisions are created according to abilities, levels of education, and training. Within these subdivisions there are further divisions established along non-economic lines, such as sex, age, race, and location of residence.

3.5 The Two Queue Theory

The first of the alternative concepts of the labour market resulted from an attempt to adopt two insights into the concept of the labour market. The first was that wages can be viewed in terms of a wage structure and market relations. Dunlop (1957, pp. 15-18) analysed wage structure in terms of "job clusters" and "wage contours". Job clusters are "stable groups of occupational classifications determined by technological conditions, administrative arrangements, and social custom." Wage rates are determined in wage contours, defined as "stable groups of firms with a specified product, type of labour required, and a geographic supply of labour." The wage contour reflects the importance of the product market. Job clusters and wage contours represent ranges of skills requirements, product markets, and geographic locations, all of which influence the size of wages. In effect, job clusters and wage contours are a means of identifying groups of buyers and sellers in the labour market.

The second insight was that the labour market does not operate perfectly. Some workers face mobility problems and show inadequacies in terms of skill prerequisite. Workers are not interchangeable. Employers express preferences for labour. Workers have preferences concerning the kind and the conditions of work. Security of existing employment may be more attractive than the net advantages that may be secured by a worker being mobile. The supply of labour depends upon the prerequisites, conditions, and the advantages of the job and work. In other words, employers seek to attain the most skilled and most experienced employee available for a particular job. The employers' hiring policies are directed according to economic considerations, such as education, test scores, and others. However, the hiring policies of employers do not necessarily only rely upon economic factors. In a study examining the hiring policies of employers, Comanor (1973) tested the hypothesis that in the traditional theory of the firm the greater are the profits of the firm, the more the firm will allocate for purchase of the more preferred factors of production. The preferred factors in this study are the more skilled and more experienced workers. The study tested and confirmed the hypothesis, but in doing so added the observation that the greater the profits of the firm, the greater is the evidence that an increase in discrimination occurs in hiring practices and policies. Therefore, the labour market not only did not operate perfectly, and non-market and non-competitive factors are necessary in an explanation of the operations of the labour market.

The two distinctions above were then elaborated and incorporated into a model of the labour market. In a presentation by Doeringer (1969), the operations of this model of the labour market were defined in terms of the <u>two queue theory</u>. In the two queue theory it is assumed that both employers and potential employees have perfect knowledge and information as to the other's standards and qualifications. The two queue theory identifies two queues that essentially establish job clusters. The first queue is the hiring queue. The hiring queue represents the demand for labour in terms of employers' hiring policies. In determining hiring policies employers order workers in large, broad groups defined not by potential productivity and wages, but by quantifiable variables, such as age, educational attainment, test scores, and by subjective interviews. The other queue is the job vacancy queue. This queue represents the preferences of the supply of labour. Employees or potential employees rank the jobs of employers in order of preference according to wages, work conditions, reputation, and other factors.

In the first instance, the labour market matches jobs and workers according to the relative positions on the respective queues. Once an employment relationship has been established, workers and employers may appraise and reappraise each other. If either party is dissatisfied, the relationship is terminated and the supply and demand queues are reassembled. Through the reappraisal process, the least acceptable and the least attractive jobs are continually rematched with the less skilled workers.

Furthermore, the two queue theory provides an explanation of unemployment as a result of the interactions of the hiring and job vacancy queues. Since the conditions of disadvantages and advantages for employment are not equal, a dividing line occurs between the employed and the unemployed. This dividing line is towards the end of the hiring queue and fluctuates with overall demand for labour. Generally, unemployment below this line is considered involuntary and a demand phenomenon. The two queue theory of the labour market has no self-regulating mechanism correcting the structural problems of unemployment.

3.6 The Dual Labour Market

In commenting upon the two queue theory, Doeringer (1969) pointed out that high labour turnover is characteristic of the employees at the bottom of the hiring queue and is largely due to the poor quality of employment indicated by the jobs being at the end of the job vacancy queue. Doeringer (1969, p. 15) maintained "that some of the demand conditions derive from the type of labour available, but other demand conditions derive from the nature of the product market, low profit levels, inelasticity of product demand with respect to quality, and low skill content of jobs." These demand conditions generated low wages, low status, and high turnover, and it is these that contributed to what appeared to be excess labour demand and involuntary unemployment. The less preferred workers at the end of the job vacancy queue adopt attitudes towards the labour market compatible with high labour turnover. Labour scarcity was the only weapon to force adjustments in pay and working conditions. According to Doeringer (1969, p. 17), if the less preferred workers were more stable, "there would be an imbalance between the behaviour of labour demand and labour supply. The jobs available to these less preferred workers are independent of labour supply, and adjustments to labour supply are made to an inelastic demand." The instability of demand for labour is transformed into voluntary turnover by the supply of labour. Therefore, we find that labour market differences exist among groups, that is the preferred and less preferred workers. In addition, as noted in the Comanor (1973) study above, characteristics, like sex and race, enter into an employer's hiring decisions alongside potential productivity. In other words, the labour market is not a competitive mechanism that evaluates workers solely in terms of economic characteristics and that eliminates the differences in the treatment of workers.

Thus, the differences and instability of workers and jobs appear to be important characteristics resulting from the operations of the labour market. The labour market can be viewed as the result of a process that encouraged market differences among groups of workers. Natural and social barriers to entry and mobility are instrumental to an understanding of the labour market and the stability of workers and jobs. Natural barriers may include physical strength, dexterity, aptitudes, academic standards, or rare innate abilities. Social barriers range from barriers against women, social groups or races to barriers against non-union members. In addition, the conditions of jobs and work tend to affect workers' preferences in the decision to work. The conditions of jobs and work, such as wages, the criteria for hiring and promotion, mobility to other firms, the type of authority and organisational structure, and the nature of the work, comprise additional labour market characteristics that may influence the operations of the labour market. The two queue theory offered an alternative frame of reference from which to consider the ideas above for an understanding and description of the labour market conditions facing groups of workers, and in particular the less preferred or disadvantaged worker. The disadvantaged worker persistently resided at the end of the hiring queue in the two queue theory and generally only found employment in the lowest paying jobs. These less preferred workers faced social barriers to employment because of their characteristics and qualifications. The disadvantaged worker may be among other things, a high school dropout, under twenty-two years of age, without work experience, over forty-five years of age with poor work habits or physically disable. In addition, some other special obstacles to employment include having a criminal record, being aged, an alcoholic, a drug addict, Black, Spanish-speaking, non-white, or female.

One result of this type of work was that Piore (1970) introduced the concept of the <u>dual labour market</u>. The dual labour market suggests that the labour market is comprised of separate sectors or segments with industrial and occupational characteristics. The dual labour market depicts the structural differences in the labour market facing the individual. In the dual labour market, workers and occupations are characterised as operating in one of two segments of the labour market, a primary segment and a secondary segment.

> The primary market offers jobs which possess several of the following traits: high wages, good working conditions, employment security and stability, equity, and due process in the administration of work rules, and chances for advancement. The secondary market has jobs, which relative to those in the primary market, are decidedly less attractive. They tend to involve low wages, poor working conditions, considerable variability of employment, harsh and often arbitrary discipline, and little opportunity for advancement.
>
> Doeringer and Piore (1971, p. 165)

On the one hand, hospitals, hotels, warehouses, maintenance service companies, industrial sweatshops - each is representative of an enterprise containing secondary segment jobs. On the other hand, engineering firms, construction companies, insurance companies, law firms, and firms of accountancy are representative of enterprises containing primary segment jobs. Often, primary and secondary jobs are contained in separate industries, but frequently examples of secondary jobs existing side by side with primary jobs are found in the industrial structure of some primary segment employers. Thus, we find that distinctions between primary and secondary jobs are not always evident. The main distinctions between the segments are the stability characteristics of the jobs and workers in the segments. The primary jobs require and develop stability, whereas jobs do not require and may even indirectly discourage stability.

The primary segment caters to workers with skills and good employment records and exhibits a very stable structure; while the secondary segment jobs require few skills and exhibit a high degree of instability. These characteristics of the segments of the dual labour market appear to be closely associated with the structure within the labour market segments, the internal labour market. According to Doeringer (1969, pp. 1 - 2),

> the internal labour market is an administrative unit within which are performed the market functions of pricing, allocating, and often training labour. It is governed by a set of administrative rules which delineate the boundaries of the internal labour market and determine its internal structure. These institutional or administrative rules define relationships between jobs for purposes of internal mobility and any privileges which occur to workers within the internal market.

Internal labour markets have geographic, occupational, and product dimensions. An internal labour market may be established by a single or multi-plant enterprise, or may be industry wide. In many circumstances the administrative units may be divided into submarkets for different categories of workers, such as maintenance, clerical, or production. The rules and procedures controlling the hiring and the internal allocation of labour distinguish between workers outside the internal market and those who have already gained access to the internal market. The rules and procedures accord to those inside the internal labour market certain privileges relative to the external labour force. The rules and regulations governing mobility within internal labour markets define the degree of job security and stability, the chances of advancement, and the extent to which due process prevails.

Three institutional and structural developments have been associated with the internal labour market: firstly, skill specificity; secondly, on-the-job training; and thirdly, customary law. Skill specificity, that is specific skill to meet the exact requirements of the job, is fundamental to many firms. Specific skill training is contrasted with general skill training that increases the marginal productivity of a worker to all firms and does not necessarily reward the firm providing the general skill training with all the economic benefits. On-the-job training provides the firm with the chance to select the best workers to train and gives them actual work experience at less than full wages. Customary law, which creates internal stability, generates the development of informal organisations within the formal organisation.

The three developments further encouraged the establishment of rules and procedures that served to impose a very rigid, self-perpetuating structure on the internal labour market. The rigidity of the rules provides a bureaucratic mechanism that is difficult to erode. In addition, technological development influenced the tendency for structural rigidity. The behaviour of unions and management under the set of internal labour market rules tends to solidify the substance of the rules in accordance with the state of employed technology. They react with caution in adapting new technological developments that may put part of the work force in jeopardy of losing jobs.

In the case of the disadvantaged worker, more times than not, he is rejected by the internal labour market in the primary segment because of his qualifications. The disadvantaged workers are customarily employed in the secondary segment. Jobs in the secondary segment do not require the skills of the primary segment workers, the more advantaged workers; therefore, employment discrimination is muted. High unemployment rates among the disadvantaged and frequent turnover are common in the secondary segment. Both are due to the dynamics of the recycling process, described by the two queue theory. The disadvantaged workers continually appraise and reappraise the limited number of employers available to them in the belief that any job is an improvement.

A major characteristic of the secondary segment is the instability generated by the jobs within it. Low wages and lack of training and promotion derive from the instability of the secondary segment. According to Doeringer and Piore (1971, pp. 169 - 175), the initial source of the instability may be attributed to many factors including:

(1) discriminatory practices which reserved stable jobs for whites and relegated jobs most sensitive to employment variability to Negroes and the disadvantaged;
(2) the legacy of agricultural practices in the Southern United States;
(3) The concentration of unstable jobs in the inner city and the movement of stable, preferred employment opportunities to the suburb.

Jobs are unstable and unstable life styles are propagated. These jobs act to reinforce life patterns antagonistic to stable employment. Regardless of the casual relationships determining instability, instability persists within the secondary segment of the labour market due to the interactions cited above. The instability of the disadvantaged and the lack of motivation on the part of employers to stabilise employment opportunities keeps the secondary segment in a state of disequilibrium.

Full responsibility for information transmittal and career and opportunity development is maintained by the operations of the labour market. The internal labour markets and the rules which govern access to or movement within them perpetuate the distinctions between primary and secondary segments and disaggregation of labour. Primary segment workers favour internal labour markets because the rigidity and the closedness of the institutional structure enhance job security, chances for promotion, equity and due process. Employers tend to rely upon internal labour markets since they reduce the costs of recruitment, screening, and training in replacement. Internal labour markets provide employers and employees with the opportunity to develop more efficient techniques of production at the expense of those potential employees outside the internal labour market.

However, the secondary labour market segment is that in which employer investment in recruiting, screening, and training is small and labour turnover is high. Many secondary segment jobs are outside any internal labour market, and labour allocation resembles the competitive operations in which jobs and workers are interchangeable. Secondary segment workers have little opportunity to express concern about security or promotion. Worker influence, interest, and motivation are difficult to develop in the secondary segment.

Ports of entry, positions through which potential employees may enter the labour market, are another aspect of the institutional structure of the dual labour market. Ports of entry are primarily at the lower levels of the occupational ladder in the primary segments. For each internal labour market there are varying degrees of openness to the external labour market through fluctuations in the number of ports of entry. Doeringer (1971) noted that approaches to the ports of entry are governed by enterprise markets or hiring hall markets. The former dominates the manufacturing and while collar jobs. The employer plays the dominant role while the union is more passive. Security is provided by the internal labour market rules. The union can negotiate to a limited degree the rules of entry and promotion but not the decisions. Hiring hall markets, central clearing houses for union controlled jobs, dominate construction and other industries where employment is of short duration. The union is the dominant factor in defining, structuring, and deciding the hiring policies. In both cases the internal labour market rules governing entry are fundamental and serve to increase the rigidity and closedness of the internal labour market structure. However, the secondary segment jobs are not usually characterised by ports of entry because most of these are dead-end jobs, that is jobs with little, if any, possibility for advancement.

3.7 Labour Market Segmentation

A further modification of the concept of the labour market has been introduced by Piore (1972). Piore took the dual labour market one step further by developing the concept of labour market segmentation or stratification. The modification was introduced to create a broader view of the labour market which not only emphasises the problems of the disadvantaged workers, but stressed equally the distinctions facing all workers in the labour force. The concept of labour market segmentation recognises the primary and secondary segments of the dual labour market. The modification elaborates upon the concept of the primary labour market segment in order to recognise the distinctions between primary jobs. Piore suggested the division of the primary segment into two separate segments; a subordinate primary segment (the lower tier), and an independent primary segment, (the upper tier). The descriptions of jobs and workers used in the discussion of the primary segment in the dual labour market are characteristic of the subordinate primary segment. In addition, the jobs and workers of the subordinate primary are characterised by routinised work, infrequent turnover and mobility, dependability, discipline and responsiveness to rules and authoritiy, and identification with the goals of the firm.

The independent primary segment is composed of professional and managerial occupations characterised by high levels of pay, status, excellent promotional opportunities, easy and smooth mobility. In the independent primary segment the mobility and turnover patterns tend to resemble those in the secondary segment, except in this case the trends are associated with advancement. Another resemblance to the secondary segment is that the independent primary segment lacks an elaborate set of work rules and formal bureaucratic procedures. In the subordinate primary segment, relationships are determined by the structure and the rules of the internal labour market; whereas in the independent primary segment an informal, internalised code of behaviour is developed. Formal, advanced professional education is a barrier to entry in to the independent primary segment. The work is characterised by great variety and opportunities for individual creativity and initiative. Piore (1972, p. 4) states "that these last characteristics differentiating the upper tier and the lower tier seem to be underlying many complaints of the middle income groups over the last several years, and the distinctions between the upper and lower tiers speak to the problems of this segment of the labour force much as the dual labour market explains those of the disadvantaged workers."

The development of the concept of labour market segmentation served to depict a labour market that not only took into consideration economic factors, but also social and institutional factors, in an explanation of the structure and operations of the labour market. In doing so, the concept of labour market segmentation identifies the dominant groups in the labour market. This socio-economic approach to the concept of the labour market further suggests the compatibility of the three labour market segments with three social class subcultures. In the following extract from Piore (1972, pp. 4-6), the secondary, subordinate primary, and the independent primary labour market segments are associated respectively, with the lower, working, and middle class subcultures. Piore wrote:

> The labour market divisions seem quite clearly related to these subcultures and possibly are, in the same way anchored in them. The characteristics of the subcultures vary over the life cycle of an individual; the parallel to the labour market segments is closest at the adult phase of the life cycle, the age when individuals have typically married and have children. The working class subculture at this stage is anchored in a stable, routinised life style. Life centres in an extended family unit and in a set of relationships in a peer group drawn from friends developed in childhood and adolescence. The individual tends to define himself and his role in terms of these relationships. Work is viewed as an instrument for obtaining the income necessary to support the family and participate in peer group activities; education is seen as an instrument for obtaining work. In all these respects, the subculture appears supportive of work in the lower tier of the primary sector, which seems like the basic life style to be stable and routinised. The priority accorded family life enables one to bear the lack of challenge on the job which might, were it to exist, distract from family activities.
>
> The working class subculture contrasts with the middle class. Here, the line between the family, on the one hand, and work and educational activities on the other, is blurred. The extended family obligations of the working class are narrowed to the nuclear unit, thus reducing the potential for conflict with work. Both work and education are viewed, at least ideally, as ends, rewarding in themselves, as well as a means for obtaining income. The friends with whom the family passes its leisure time are often drawn from work and based upon common professional interests. In these respects, the middle class subculture is well adapted to the support of upper tier work patterns; the nuclear family and professional friendships facilitate geographic and social mobility and permit intellectually demanding and time-consuming jobs. The view of education is supportive of extensive prework schooling far removed from the payoff and of no immediate relevance.
>
> The lower class subculture deviates from that of the working class in a way which appears similarly adapted to employment patterns of the secondary labour market. Lower class men have a highly personalised conception of themselves, divorced from an independent of a network of relationships with family and friends. Such relationships thus tend to be volatile, short-lived, and unstable, and their life tends to be characterised by an effort to escape routine through action and adventure. It is a pattern consistent with the erratic employment of the secondary labour market as well as with other characteristics such as the personal relationships between worker and supervisor.

The line of discussion in Piore's work does not end here, but proceeds to analyse labour market segmentation in terms of the dynamic process which determines it. The labour market segments reflect differences in what Piore (1972, p. 6) calls <u>mobility chains</u>, a notion that "movement in society is not random, but tends to occur in more or less regular channels." The labour market segments can be concluded as embodying types of job sequences through which an individual passes. The distribution of workers between the segments is dependent upon the mobility chain through which the workers pass. Piore (1972, p. 8) remarks "that the critical distinction between the primary (the upper and lower tiers) and the secondary sectors is that the mobility chains of the former constitute some kind of career ladder along which there is progress toward higher paying and higher esteem jobs." To proceed any further will exceed the limited purposes in this chapter of presenting the development of a concept of the labour market that embodies social, institutional, and economic factors in its structure. This brief presentation of the relations to class subcultures and mobility chains serves to establish the socio-economic nature of the concept of labour market segmentation.

3.8 Summary

In this chapter the purpose has been to establish the importance of a concept of the labour market that reflects the social, institutional, and economic environment. We presented the conventional approach to the labour market as a starting point from which to establish modifications and to indicate the general direction required to develop an alternative structural framework of the labour market. We recognised that the division of labour promotes occupational specialisation. This and other factors (e.g. social, ethnic, sexual), and internal labour markets, act as non-market and non-competitive factors that influence the structure and operations of the labour market. The development of non-competitive models, such as the two queue model and the dual labour market, has led to the concept of labour market segmentation.

The notion of labour market segmentation recognises a threefold division of the labour market. These divisions closely resemble the lower, working, and the middle classes of the social structure. The hierarchical classifications of labour market segmentation are consistent with the hierarchical tendency in the distribution of personal income in the United States pointed out by Reder (1969). Reder observed that the degree of respectibility of an individual or family varied with the share of income received by the individual or family. Family heads normally receive a larger share of income than the other members of the family. Important families receive a larger share of the communities income than less important families. In labour market segmentation, important jobs receive higher pay.

In conclusion, as well as stating the conditions and the pattern of development of groups in the working population, the concept of labour market segmentation exhibits the three factors in operation in the labour market, noted by Dunlop (1950), that bear an influence on the pricing and the allocation of the market factors. Labour market segmentation disaggregates the economic structure into three classes or segments within which industrial and occupational characteristics determine how buyers and sellers cluster. Entry and exit are determined by the internal labour market structure. It is the purpose of the subsequent chapters in this thesis to identify the groups of labour associated with each of the three labour market segments described in the concept of labour market segmentation.

CHAPTER 4

Conceptual and Technical Considerations for Labour Market Segmentation

4.1 Classes of Labour

As mentioned in Chapter 3, the development of an alternative and workable framework for labour market analysis requires the identification of dominant groups of similar participants in the labour market. In this respect, we regard labour market analysis as relevant not only to the problem of equity, but also to the relations between the buyers and sellers in the labour market. In order to develop any hypotheses from an alternative framework of the labour market, it is necessary to make a study of the factors that underlie the labour market structure and encourage the formation of groups of workers. A study of these factors will serve to indicate the important influences on an individual's position in the labour market, and, in turn, his position in the earnings structure.

However, in trying to identify the dominant groups (or segments) in the working population, an overall distribution of personal income provides only the means for a crude first approach. We saw in Chapter 2 that the theoretical-statistical approach to the distribution of personal income generally examines labour in the aggregate, implicitly assuming away the problems of analysing the quite distinct subgroups of labour. The sociological approach analysed a number of factors contributing to differences in the earning power of individuals, but not in terms of distinct subgroups. Even the traditional labour market analysis presented in Chapter 3, although recognising different types of labour, assumes they are otherwise homogeneous, in the sense that labour is perfectly mobile, has perfect foresight, and seeks to maximise the trade-off between work and leisure. In other words, in conventional economic analysis of both the distribution of personal income and the earnings process, differences in labour were reflected in differing levels of income, but despite these differences labour still constituted one aggregated class.

By contrast, the presumption in this thesis is that the disaggregation of one ostensibly homogeneous class of labour into a number of separate, homogeneous classes increases our knowledge of the allocation and pricing of labour in the labour market. This provides a framework for associating motivations and behaviour to the factors that influence the structure of the labour market. We regard the disaggregation of labour into separate classes as a necessary first step to an understanding of the entire labour market. We are not fragmenting for the sake of fragmenting, but to perceive the labour market as a pattern of relations between its parts.

Of course, the use of classes of labour to analyse the earnings process is not a novel approach. Ginsberg (1929, p. 554) pointed out that "the idea of classes has played a major role in social theory and political movement, but very few scientific studies exist into the nature of class differentiation and its conditions." In a short summary Kerr (1957a) dated the recognition of classes in economic thought back to the early Classical theorists who attempted to establish the link between the social, institutional, and behavioural characteristics of the economic system. Firstly, there were the natural class theorists, such as Smith, Ricardo, and Malthus, who maintained the existence of natural laws at work establishing classes and income shares. Secondly, the social class theorists led by Marx defined social laws relating the class structure of society to the fundamental laws or forces fixing the factor shares in the economy. Thirdly, the natural individual theorists believed the physical relations governing supply and demand determined income distribution. Fourthly, the social group theorists challenged this view with the belief that human institutions governed distribution. Finally, the socio-economic theorists as described by Gendarme (1968) analysed the income distribution process in terms of the influences and impacts of social groups in the working population, their conflicts, and historical, ideological, and political factors.

A moment's reflection on the economic system in the United States may suggest there is not a single, large homogeneous class of labour. Different types of labour exhibit different motivations; and there is great variety in the manner in which labour is allocated and awarded an income. Kerr (1957b, p. 173) noted a similar observation by Smith "that it is a natural state of things in that workers, being more diverse and numerous, have less a community of interest than entrepreneurs and the need for more and different kinds of bonds." Likewise, Whitaker (1974) argued that Marshall, following Mill and Cairnes, recognised a social hierarchy of grades of labour distinguished by different trades or occupational groups, that is non-competing groups, and that restrictions between grades inhibited the mobility of labour.

The purpose of disaggregating labour and of specifying homogeneous classes of labour is to arrange together those individuals which are most alike and to separate those most unalike. The framework of the labour market (the structure creating the classes of labour) determines which individuals lie in which classes. The individuals do not determine the classes. Something more than mere resemblance is meant by the classification of individuals into groups. A bond or relationship between a group and its members may be revealed by our classifications. The descriptive characteristics of the individuals may be used to reveal any relationships between a group and its members. Furthermore, the degree of concurrence, or the correlation, between characteristics might indicate general patterns of association, even though classes of labour are not composed of perfectly identical individuals. In certain cases the members of a class might differ markedly from each other with respect to certain characteristics, yet they are classed together. Some underlying association of characteristics might reveal the bond linking the members. Consequently, a primary objective of this thesis is to identify the important characteristics and associations underlying the structure of the classes of labour.

In Chapter 3 the discussion of the labour market hypothesized the compatibility of the development of the concept of labour market segmentation with the industrial and economic development of the United States. In addition, the discussion suggested the applicability of the concept of labour market segmentation to an analysis of the earnings process and to the determination of classes of labour with characteristic features. Therefore, in this thesis we adopt Piore's (1972) concept of labour market segmentation as the framework for specifying classes of labour. Consequently, we will present the concept of labour market segmentation in a form suitable for analysing the composition and the associations within and between the classes of labour (labour market segments). The concept of labour market segmentation, as discussed in Chapter 3, describes the operations of the labour market in terms of economic, and non-market and non-competitive forces. For example, under economic forces we can include the industrial composition, the unemployment rate, the pattern of work (hours and weeks worked), the earnings structure, and others. Under non-market and non-competitive forces we can examine institutional barriers to mobility, family and social structure, geographic residence, transportation, and others. We will concentrate upon these economic, and non-market and non-competitive factors in the belief that these factors genuinely influence, if not determine, the segments. This is in contrast to the view that labour market segmentation and the accompanying classes of labour occur by chance. To be more precise, we assume that the conditions and operations recognised as part of the concept of labour market segmentation create the separate, homogeneous classes, or segments, in the labour market and determine the composition of these segments.

4.2 Technical Considerations

In the section above, we referred to Ginsberg's (1929, p. 554) observation that "the idea of classes has played a major role in social theory and political movement, but very few scientific studies exist into the nature of class differentiation and its conditions." Social theory and economic theory have recognised the importance of the actions and the motivations of groups within society. However, so it is today as it was when Ginsberg (1929, p. 554) further noted that "either because of an absence of adequate classification or a lack of agreement with regard to criteria of classes of labour," social scientists and economists have often denied the existence of specific cohesive classes.

In the past, the recognition of groups in the labour market had been suggested, but usually these groups were merely the dominant income size classes observed along the spectrum of the distribution of personal income. Conventional labour market analysis did not provide or fully consider specifying lasting and specific classes of labour. However, the introduction of the concept of labour market segmentation has offered the possibility for providing a framework of classification. The concept of labour market segmentation is not to be regarded as a purpose in itself, but as a mechanism for changing our way of looking at the earnings process and a means for providing an alternative direction in our attempts to understand and attack the earnings process. In this thesis we will elaborate the concept of labour market segmentation in terms of the characteristics and the associations of characteristics of the groups of labour in the labour market segments.

We reject the use of a linear scale of income size in the specification of separate homogeneous or cohesive groups of labour because so many social, institutional, and economic factors are overlapping and are ignored. The concept of labour market segmentation proceeds beyond one-dimensional income size segments to multi-dimensional generalised segments. The emphasis is upon the formation of labour market segments and the identification of classes of labour according to a multi-dimensional or multi-factor scale.

Just as the institutions and the socio-economic conditions vary and change in importance over time, so has the emphasis as how to view labour and classes of labour. The definition of the classes of labour depends upon the purpose for which the definition is to be used. Emphasis upon certain aspects of any definition may yield different pictures of the motivations and the behaviour of labour. With changes in the definition of classes of labour over time, we must recognise that each definition is valid in its own purpose. At best, the changing views with respect to labour in economic analysis may provide the basis for the presentation and the enunciation of hypotheses to suggest directions for further research and enquiry.

Use of the concept of labour market segmentation reflects a new attitude on the part of the academic economists to the structure and operation of the labour market. The concept of labour market segmentation recognises the motivational patterns of groups and institutions, as well as economic forces, in the earnings process. The framework of labour market segmentation enables the possibility for generating and accumulating statistical descriptions to reflect the changing economic thinking. The elaboration of labour market segmentation in this thesis is an attempt to present this framework in a statistically workable and useful form.

Unfortunately, the data have been determined by institutions in society for varied purposes; and hence, the accumulation of data does not necessarily accommodate changes in the standard categories employed by economists. The introduction of labour market segmentation into economic thought did not and was not accompanied by a data collection source that permitted the specific identification of the groups of workers in the labour market segments. None of the data available from the various sources is specifically collected to accommodate the concept of labour market segmentation. Although labour market segmentation is an attractive way to organise experience in the labour market, it does not lend easily to empirical testing because of the lack of specifically organised labour market data.

A major task of this thesis is to transform the existing data into a form compatible and consistent with the concept of labour market segmentation. The task involves the reorganisation of a large body of data to meet our specific needs. To this mass of data we must apply statistical procedures and methods and develop informal and formal rules for the estimation and the interpretation of the information pertaining to labour market segmentation. The application of statistical procedures and methods in the service of quantifying labour market segmentation does not guarantee the specification of the most ideal segments but may offer an insight into the structure and operation of the labour market.

Before discussing the technical considerations involved in transforming the available data into a form consistent with labour market segmentation, we find that an appropriate starting point is a brief restatement of the concept of labour market segmentation. Labour market segmentation is an alternative model of the labour market in which labour demand is divided into three segments. Each segment is distinguished by different labour market characteristics and rules.

The concept of labour market segmentation identifies a hierarchy of the occupational structure contained within different segments of the labour market. The segments are horizontal sections of workers and jobs at about the same level in the vertical occupational hierarchy. The hierarchy can be considered an occupational ladder with the least attractive occupations at the bottom and the most attractive occupations at the top. Following Piore (1972),[1] the concept of labour market segmentation that we adopt identifies three segments: firstly, the secondary segment at the bottom of the occupational ladder; secondly, the subordinate primary in the middle; and thirdly, the independent primary segment at the top of the occupational ladder. The further up

---

1) For a more detailed presentation of labour market segmentation refer to section 3.7 of Chapter 3 above.

the occupational ladder an occupation is, the more specific and specialised are the job requirements. As a candidate attempts to progress up the hierarchy, the requirements become progressively more difficult to fulfill, and entry becomes more limited as interviewing and screening procedures intensify.

The secondary segment is composed of the less attractive occupations involving low wages, poor working conditions, considerable variability in employment, harsh and often arbitrary discipline, and little opportunity for advancement. The workers in the secondary segment are generally unskilled, less educated, and often members of minority groups. The subordinate primary segment offers jobs which are characterised by high wages, good working conditions, employment stability, job security, and routinised work. The workers in the subordinate primary experience infrequent mobility, are responsive to rules and authority, and identify with the goals of the firm. The occupations in the subordinate primary are more skilled than those in the secondary segment. These occupations are the white collar jobs, the skilled trades and crafts, and lower and middle level management positions. The independent primary segment is composed of professional and upper level managerial occupations characterised by high levels of pay, status, excellent promotion opportunities, and easy and smooth mobility. Formal advanced professional education is normally a requirement of entry into the independent primary segment. The work generally offers opportunities for creativity, initiative, and variety.

To achieve a consistent fit of the data to the definitions of the labour market segments, we must make several assumptions about the labour market segments. In this way we will arrive at an ideal specification of the labour market segments. In practice the ideal is often unworkable:

a minimal divergence from the ideal may be the only achieveable goal.

Ideally, the first assumption we must make is that the labour market segments are mutually exclusive. This implies that the range of individuals employed in an occupation must fit exclusively into one of the three labour market segments. In the real world, the labour market segments could cut across some of the occupations, but in those cases the responsibilities and qualifications of the individuals would be different. These occupations may encompass different grades, and some grades may be in the independent primary segment, whilst some grades may be in another segment. For example, some accountants doing routine clerical work may be in the subordinate primary segment; whilst other accounts (tax experts) may be in the independent primary segment. In effect, if data were available, we could redefine each grade of such occupations as a different occupation and fit each into one of the segments. In this case, each grade would be considered a separate occupation. Nevertheless, the data are not available, and we can only assume that bonds of association may exist that in the end influence the grouping of all the grades in the same labour market segment. Thus, we are forced to ignore the possible existence of grades of occupations. With this assumption of mutual exclusiveness, we intend to minimise the overlap between the labour market segments.

A second assumption is that from the viewpoint of the employers, workers within a labour market segment are considered homogeneous in terms of the wage being offered. An employer would offer the same wage for each member within a labour market segment. In reality, this is not the case, wage differentials may be offered along non-market and non-competitive lines, but by initially considering the workers homogeneous, we have a cohesive group from which to analyse the associations between the wage and the economic, and non-market and non-competitive factors.

A third assumption is a low degree of mobility from one labour market segment to another. The introduction of this assumption enables us to highlight the significant characteristics of the segments that generate patterns of labour allocation to the segments. Two more assumptions, a fourth and fifth, are the stability of the labour market segments through successive periods and across geographic boundaries. These two assumptions imply that the segments exhibit constancy throughout the economic system. To close this section we must state that the main purpose of listing these assumptions is to define the <u>ideal</u> boundaries of the concept of labour market segmentation. However, regardless of the assumptions concerning the specification of labour market segments, we must emphasise the inability to specify perfectly homogeneous labour market segments.

4.3 Conceptual Considerations

In the past, studies have examined groups in the labour market and in the distribution process. Initially the groups were arranged by income size, by industrial sector, and by occupation according to income. Now, the concept of labour market segmentation and the statistical tools available facilitate the groupings by occupation according to a multi-dimensional scale. We will use the economic, and non-market and non-competitive characteristics of labour to determine the labour market segments. The focus is on the ability of the individual to earn income under the social, institutional, and economic conditions of the labour market. The form in which labour is supplied (viz. the occupation) reflects imprecisely the level of educational attainment, age, skill, desire for consumption, and other personal characteristics of the individuals in an occupation. The demand for labour determines the participation rates and the skills required.

By concentrating upon the characteristics of the work force and the factors influencing it in the determination of the groups in the labour market segments, we are emphasising the importance of the occupation and the occupational structure in the labour market and in the economic system. Occupational groupings define groups which are homogeneous with respect to chance influences acting upon individuals, such as equality of opportunity, ability, inheritance. An individual's perception of effort, reward, values, and beliefs in the cultural, social, institutional, and economic environment is reflected in the occupational choice. Just as the entrepreneur is essential to production and distribution in that he brings capital and labour together in the production process, so equally essential is the form in which labour comes into contact with the machinery of production, namely, the occupation. The occupation consumes more of an individual's time than any other element and is one of the more dominant influences in an individual's life. For the individual, Miller (1955, p. 49) describes the occupation as "the pivotal element in the production process and the distribution of the national output."

This is an appropriate point for us to discuss briefly the concept of labour market segmentation, based as it is on the occupation and occupational categories, with respect to the unequal distribution of personal income. Any framework for understanding the earnings process or the distribution process should incorporate as part of its structure the factors that influence the differences in earnings or the degree of inequality. Kravis (1960, p. 411) specified a list of factors or conditions that should be incorporated into an analysis of the earnings process or distribution process. Kravis' list identifies four types of factors that influence the degree of inequality:

1. Distributions of human characteristics affecting ecnomic performance (motivation, education, ability, age, chance).
2. Barriers to economic mobility imposed by the state or embedded in the mores of people (legislative, racial or social discrimination).
3. Economic structure which may or may not provide for highly differentiated forms of work with large income differentials.
4. Social and political organisation.

Specifying the labour market segments according to occupational groupings complies with this list of Kravis'. Since the labour market segments are to be determined in part according to economic factors, such as the industrial structure, employment rate, education, location of employment, income structure, and others, we can assume that items 1 and 3 are taken into consideration in our analysis. Similarly, since the segments are also to be determined in part according to non-market and non-competitive factors such as barriers to mobility (sexual and racial), family structure, residence, and others, we can further assume that items 2 and 4 in the list are included in our analysis.

The occupa tion, and in turn, the labour market segments, reflect the motives and the scope for action of classes of workers in the process of the determination of personal income shares. Labour market segmentation positions classes of workers in the earnings process, and it is therein that the labour market segments influence the distribution of personal income. In short, the placement of occupations into labour market segments may indicate the contribution to the national product and the influence upon the nature of the distribution curve of homogenous groups of workers. However, these implications of labour market segmentation for income distribution are beyond the scope of this thesis.

Nevertheless, the specification of the groups in the labour market segments, as we will determine them, reflects the four items in Kravis' list that influence inequality. However, in the context of this thesis we regard these four items as more appropriately influencing the differences in the allocation and pricing of labour in the labour market.

4.4 Summary

We will determine the groups of labour described by the concept of labour market segmentation in terms of a multi-dimensional scale of economic, and non-market and non-competitive factors. These factors will be measured on occupations. Even though we allocate occupations into the labour market segments, we do not pretend that this will enable us to explain the diverse behaviour of labour markets throughout the economy. However, the framework may provide a new view of existing circumstances. The view may not be correct, but even so, it may provide fresh insights into the conventional framework.

To be realistic, we must admit that in reducing labour to significantly large and manageable groups, we may sacrifice homogeneity within these groups. It follows that we may possibly weaken the link between the macroeconomic and microeconomic aspects of the labour market. But, the advantage of the approach we adopt is that it considers a broad range of factors which may determine the labour market segment within which any worker lies. In conclusion, we assume that the specification of the dominant groups in the labour market according to the concept of labour market segmentation will increase our understanding of the earnings structure and will indicate the connections between the earnings structure as a whole and the factors determining an individual's position in this labour market structure.

CHAPTER 5

Data Description

5.1 Data Considerations

For the empirical work a careful selection of the data is important, and the form of the data to be chosen should fit as nearly as possible the purpose for which it is to be used. In most studies of the distribution of personal income and of the labour market, the important issue is how people share the final distribution of the command over the goods and the services produced by society. But, in studying this important issue there are differing underlying purposes. In the first place, producers may study the distribution to estimate the probably extent of markets in order to establish production processes and quotas. In the second place, policy makers may analyse the earnings process and distribution in order to determine the structure and the magnitude of taxation. In this thesis, the purpose of studying the labour market is to determine the important factors operating in the labour market that influence an individual's occupational choice, position in the labour market, and, in turn, ability to earn income. The occupation, an individual's source of income, is fundamental in determining an individual's position in the earnings process. Our purpose is not to pass judgement on the earnings process and the distribution of personal income, but to seek the probable labour market structure and factors that place individuals into certain labour market groups.

In Chapter 4 we stated that the occupation of an individual would provide the starting point for our analysis of labour market segments. The two most important requirements for the selection of the occupational data are availability and consistency. Before presenting the data source, the elementary considerations of the data relevant to the analysis are discussed. Our discussion of data considerations reveals the type of organisation of the occupational data required.

The first data consideration involves the choice of the definition of the recipient unit. The recipient unit, the basic unit of the data, may be either of two forms; (1) the family or (2) the individual. The choice of the recipient unit depends upon the purpose of the study. If the study were to concentrate upon inequality and its welfare issues and implications, the use of the family unit and family income would be appropriate. The reason for this is that even though incomes accrue to individuals via the occupation, individual consumption patterns do not project an accurate consumption pattern in the economy because the family acts as a unit in the consumption process. In this sense, then, the family income bears most heavily on the future course of the economy. In other words, the family unit and the family income are useful to study the implications for the economy as a whole. The family income and the family status may determine how many and what form individuals in the family need to be income recipients. Family consumption patterns and the number of income recipients per family are important aspects of the distribution of family income. These are issues of equity and welfare relevant to the distribution of family personal income but are beyond the limited scope of this thesis.

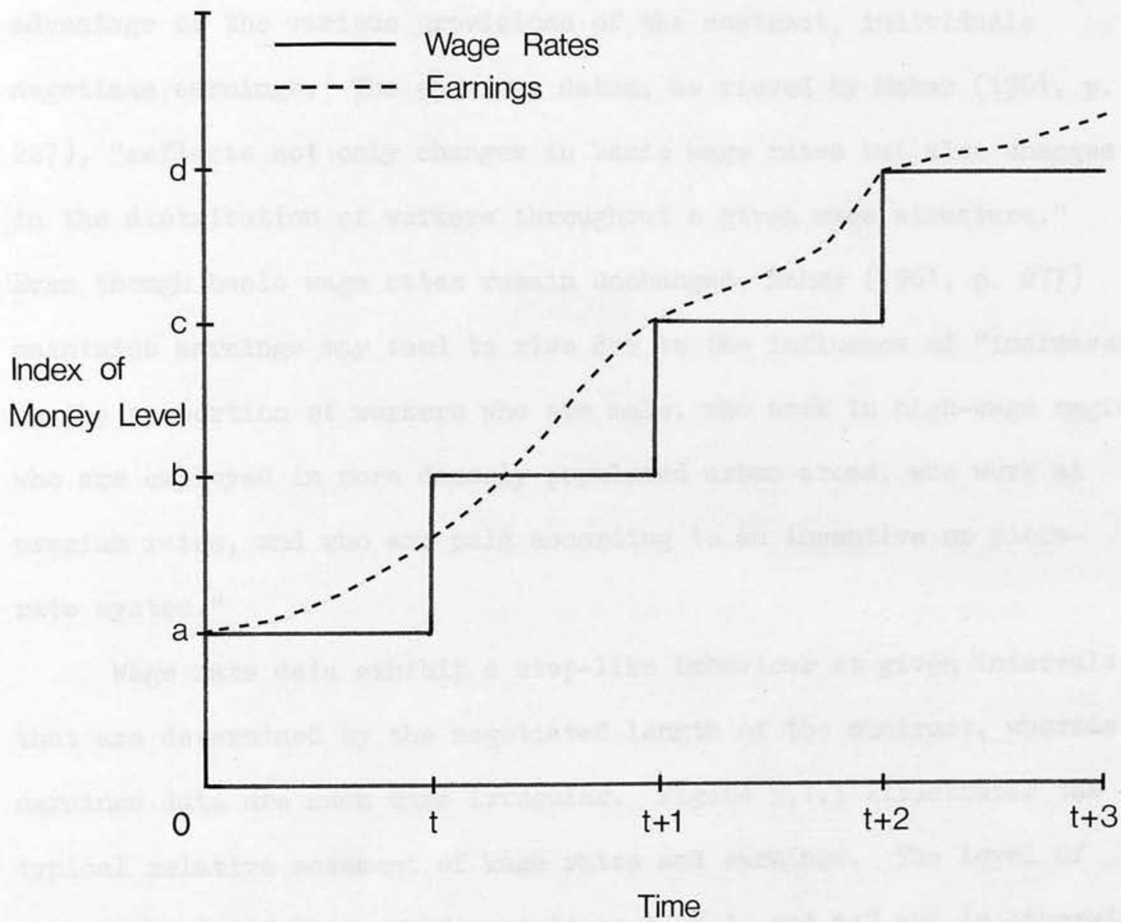
The purpose of this thesis is not to pass judgement on equity and welfare, but to examine the earnings process in the context of labour market segmentation. Thus, the causes of an individual's position in the distribution of personal income are sought within the constraints of the labour market. It is the individual who acts and is acted upon within the labour market structure in making the basic decision of occupational choice. The individual income recipient is the relevant data unit to illuminate the factors that influence an individual's scope for action within the structure of labour market segmentation.

The obvious reason for choosing the individual as the recipient unit is to discuss the probable causes that put individuals into certain labour market groups. The use of data on individuals enables us to achieve the goal of presenting a multi-dimensional analysis with which to determine the groups in the labour market segments. The basis for the multi-dimensional analysis is the occupation held by the individual. Generally, the use of the family as the recipient unit automatically assumes the conventional roles associated with husband and wife, the nucleus of the family unit. Formerly, the fact that a wife worked indicated that the husband's income was not sufficient to sustain his family. The wife's income filled the gap. Nowadays, the fact that a woman works is more an indication of the increasing independence of the wife than an indication of the wife's necessity to make up the difference between male earnings and family expenditure. Therefore, another reason for choosing the individual as the recipient unit is that it enables us to recognise the opportunity of both men and women to participate in a productive economic activity. For these reasons the recipient unit chosen is the individual.

There are other aspects of the data worth considering. One aspect of this consideration is that data on income are of two forms: wage rates, and earnings. According to McCormick (1969, p. 105), "it is suggested that earnings reflect the workings of competitive forces whereas wage rates are the result of institutional pressures." Wage rates generally are negotiated by unions and are an indication of the

Figure 5.1.1

The Movement of Wage Rates and Earnings



interactions and the attitudes of the bargaining units in the labour market. The wage rate datum is a measure of the price of labour, which is fixed at a basic contract rate. However, the contract negotiated by the bargaining units in the labour market contains provisions, such as piece-rate allowances, incentives, overtime rates, bonuses, shift allowances, by which an individual may increase income above that computed from the basic contract wage rate. By taking advantage of the various provisions of the contract, individuals negotiate earnings. The earnings datum, as viewed by Maher (1961, p. 227), "reflects not only changes in basic wage rates but also changes in the distribution of workers throughout a given wage structure." Even though basic wage rates remain unchanged, Maher (1961, p. 277) maintains earnings may tend to rise due to the influence of "increases in the proportion of workers who are male, who work in high-wage regions, who are employed in more densely populated urban areas, who work at premium rates, and who are paid according to an incentive or piece-rate system."

Wage rate data exhibit a step-like behaviour at given intervals that are determined by the negotiated length of the contract, whereas earnings data are much more irregular. Figure 5.1.1 illustrates the typical relative movement of wage rates and earnings. The level of wages, the solid line, shifts at times t, t+1, and t+2 and is otherwise constant. The level of earnings, the dotted line, shifts at irregular intervals or points which may exhibit a drift upward or parallel with the level of wages.

In Figure 5.1.1 earnings at some points cut below the wage rate line. This is possible because during the periods of lay-offs, illness, short weeks, and technical penalties and fines, earnings are below the

the wage rate. The influence of piece-rates, overtime, and others may tend to raise earnings above the wage rate at certain points in Figure 5.1.1. The earnings data can be adjusted to consider the significance of the factors that contribute to earnings being below or above the wage rate, if necessary. However, the important point of emphasis is that earnings data reflect these influences, whereas wage rate data do not necessarily. Earnings data reflect the relative results of an individual's economic activity in a period, and thus the total contribution of an individual to the productive process in the labour market. Therefore, we regard earnings data as consistent with the purpose of this thesis and select earnings data as the form of income.

The choice between before-tax and after-tax income should also be considered. Before-tax income data are the money incomes from the participation in a current economic activity, while after-tax income represents the effects of the redistribution of incomes. After-tax income data are an indication of the effectiveness of a tax structure to reduce or to alter the outcome of the earnings process and the distribution of personal income. Because we regard before-tax income data as more appropriately representing the basic return to labour for a productive activity than after-tax income data, we prefer to use before-tax income data in this thesis.

Note that up to now, we have been discussing money income; however, money income need not represent total employee compensation. The employer, in determining income, may take the tax structure into consideration. Various forms of income in kind and deferred income may be provided by the employer. These are not included in the income data on the individual and his earnings. According to Garvy (1954), deferred income appears where an employer contributes to group health insurances, social security, private pension funds, and, in special cases, stock investment programs for employees. Deferred income is an important element in the changing composition of income but data on it are not readily available.

The last aspect of the data we will consider is the spatial aspect. The major area of study should be the United States; but because of the enormous amount of data that are available on the fifty states of the United States, we prefer to limit the area of study to more manageable proportions. To do so, we limit the area of study to one of the Federally recognised regions of the United States, namely, the New England Region. A region has neither political nor economic sovereignty. However, the Federal departments and agencies administer Federal programs according to the recognised regional boundaries. A region is composed of contiguous states that are subject to similar economic influences, social customs, and geographic influences. We have chosen the New England Region which consists of six states: Connecticut, Rhode Island, Massachusetts, Vermont, New Hampshire, and Maine. The New England Region was selected rather than an individual state because of the nature of the data used in the thesis, (To be discussed in the next section). Using a region provided a manageable sample on all the occupations; whereas using one state may not have provided a sufficiently large sample for representing all the occupational classifications.

5.2 Data Source - Data Collection, Sample Design, Accuracy

The empirical work of this thesis is based on a "public use sample" created from the 1970 United States Census of the Population data. A public use sample is a sample of individual persons and households taken from the 1970 Census. The entire 1970 Census data presented in the available published Census volumes was not selected because the data on occupations were presented at different and inconsistent levels of detail for different categories of individuals in the population. In the 1970 Census the list of occupations used for some tabulations on all males usually contained as many as 442 occupations; whereas for other tabulations, the list of occupations on males and subcategories of males moved at times to as few as 170 occupations. In addition, the list of occupations used for tabulations on all females contained 103 occupations. Furthermore, the tabulations created according to occupation often presented the totals on different characteristics, such as race or sex, by different ranges of values. For example, the tabulation of occupation by industry for all males was presented with different industrial breakdowns than those for all females. Thus, the consistency and the comparability of data for different characteristics, such as sex and race, was hindered by the presentation of the data in the available published Census volumes.

With the available published Census volumes we would only have been capable of presenting separate analyses of the labour market conditions facing the different types of subpopulations. However,we wanted to present an analysis of the labour market for the total population in the labour force. It was not the case that the 1970 Census of the Population did not contain complete and consistent data at the same level of detail, but the extra expense and work involved to produce this type of published data for the use of the general public was not warranted by a corresponding demand by the public. Nevertheless, the U.S. Department of Commerce provides access to public use samples of the Census basic records. From the public use sample the individual users can make their own tabulations of the data to suit their requirements.

As described by the U.S. Department of Commerce (1971), a public use sample contains the responses to the Census questionnaire for a statistically selected sample of the households in the United States. The data are not available in the form of published volumes, but in the form of computer tape files. No detailed personal or geographic information appears in the sample. But, the public use samples provide detailed information on the demographic characteristics of the individuals. These data are not appropriate to study small geographic areas because the geographic areas are defined to be over 250,000 persons to avoid disclosing detailed information for specific individuals.

The 1970 United States Census of the Population was conducted primarily through self-enumeration. To every household in the United States a census questionnaire was delivered by mail around April 1, 1970. Three-fifths of the households (mail areas) were asked to fill out and return the questionnaire by mail. The questionnaires were reviewed for incomplete or inconsistent information. If inconsistent or incomplete information occurred, either a telephone follow-up or a personal visit follow-up was made. The other two-fifths of the households (non-mail areas) were asked to complete the questionnaire and return it to a Census official making a personal visit to collect it. If a questionnaire contained inconsistent or incomplete information, the official would complete it by interview during the visit.

There were three types of census questionnaires used for the 1970 Census. Eighty percent of the households answered the standard census questionnaire containing a limited number of standard census questions; while the other twenty percent completed one of two types of questionnaires. The two other questionnaires contained the same questions as the standard census questionnaire, as well as a number of additional questions specific to each of the two other questionnaires.

For the collection of the 1970 census data the U.S. Department of Commerce (1972b) used a random sampling procedure to determine which of the three questionnaires a household received. If a unit were designated in advance of the census as a group quarter, every person in the group quarter was considered a separate sampling unit; otherwise, the housing unit was.[1] In both non-mail and mail areas, address registers of household units were drawn up. In each area a random line was specified as the starting point for selecting the sample. The household unit on every fifth line after it was selected to comprise the twenty percent sample that would complete the two other non-standard census questionnaires. If a unit was a group quarter, all persons in the unit were placed separately on the address register, and every fifth person was selected for the twenty percent sample.

The twenty percent sample was subdivided into a fifteen percent and a five percent sample. Every fourth member of the twenty percent sample was selected for the five percent sample. The remaining members of the twenty percent sample became the fifteen percent sample. The members of the five percent sample completed one of the two other census questionnaires, and the members of the fifteen percent the other. Some questions, such as those of the standard census questionnaire answered by eighty percent of the total population, were common to both

---

1) A group quarter is a living arrangement for other than normal, ordinary family household life. Such quarters are generally found in dormitories or in houses or apartments uses as rooming houses or as partnerships in residence, if the occupants are unrelated to the owner or primary tenant.

the five percent and the fifteen percent questionnaires; while other questions were specific to either the five percent or the fifteen percent questionnaires. According to U.S. Department of Commerce (1972b, p. 194) estimates for the United States as a whole, 19.6 percent of the total population and 19.7 percent of the total housing units answered either the five percent or the fifteen percent questionnaires.

With the 1970 Census samples estimates are made for the sample populations that would have resulted had the sample been collected on a population stratified into the relevant groups in the population. A weighting procedure was used to adjust the sample data to represent these relevant groups. For each group of persons within a geographic area a ratio estimation procedure was performed to estimate a representative population.[2] According to the U.S. Department of Commerce (1972b, p. 197), "the estimates for a sample area are, in general, consistent with the complete count for the population and housing unit groups used in the estimation procedure."

The public use sample in this thesis is a one-in-a-hundred sample drawn from the fifteen percent sample census data for New England. The one-in-a-hundred samples were selected to represent "each geographic area and stratum in proportion to its frequency in the population." (U.S. Department of Commerce, 1972b, p. 197). Each member of the fifteen percent sample was given a sample weight and on the basis of the weight was classified into one of the strata. A unit was selected for the one-in-a-hundred sample if the addition of its sample weight caused the total sample weight of the stratum to exceed a multiple of

2) For further information on the ratio estimation procedure see U.S. Department of Commerce (1972b, Chapter 5).

100. If an occupied housing unit was selected, each member of the household became part of the one-in-a-hundred sample. In other words, each one-in-a-hundred sample is a stratified subsample of the fifteen percent sample.[3)]

Any comparison between census publications of the 1970 census data and the estimates based on one-in-a-hundred samples from the five percent, fifteen percent, and twenty percent samples need not be in exact agreement. The estimates are subject to variability due to the selection of the population for the public use samples. The degree of variability is described by the U.S. Department of Commerce (1973, Appendix C, p. 8) as follows:

> The chances are about 2 out of 3 that the difference (due to sampling variability) between the sample estimate and the figure that would have been obtained from a complete count of the population is less than the standard error. The chances are about 19 out of 20 that the difference is less than twice the standard error and about 99 out of 100 that it is less than $2\frac{1}{2}$ times the standard error.

However, non-sampling errors must also be recognised in the 1970 census sample data. The most obvious source of non-sampling error may have occurred when the 1970 census questionnaires were checked for incomplete and inconsistent information. If incomplete or inconsistent information occurred, new allocations were often made based on the known characteristics from the questionnaire. In other cases, the allocation was derived from a questionnaire of a person with similar characteristics or an allocation was merely designated as inappropriate. The editing and allocation procedures were designated to reflect local geographic characteristics. Nevertheless, editing procedures, no matter how sophisticated, cannot detect all the possible inconsistencies and cannot duplicate perfectly the actual responses of persons for all

---

3) For further information on the selection of the one-in-a-hundred samples see U.S. Department of Commerce (1972b, Chapter 5).

the incomplete and inconsistent information.

Other sources of non-sampling error may exist in the 1970 census data. A non-sampling error in earnings data may occur because earnings are reported from memory, not actual records, and are subject to memory bias. The census data may overstate the skill level while understating the earnings level. An understatement of employed persons is a possibility due to the omission of marginal workers, such as students and housewives, by the respondent. Despite the possibilities for error, the 1970 Census data are considered to be collected, sampled, and processed accurately and consistently with the minimum of human and mechanical errors.

5.3 <u>Data Preparation</u>

As the data was not in a form suitable for an analysis of labour market segmentation, we needed to transform the data into an appropriate form. Therefore, we had to develop a systematic procedure for transforming the data. Our general intention in this section is to provide an overview of the work involved in preparing the data for the analysis. This procedure involved three processing stages. Firstly, since there were irrelevant data on the files, we had to examine the data items and select from the data only those items relevant to our analysis. Secondly, we needed to develop a computer data processing procedure that would allow us to obtain the relevant data items and then put them into a form appropriate for an analysis of labour market segmentation. Thirdly, to ensure that the data was processes accurately, we had to check the data at each step of the computer data processing procedure.

In the first stage we found from the 1970 United States Census of the Population fifteen percent sample that three one-in-a-hundred samples were drawn, one each for the three following geographic areas:

(1) country groups; (2) states, and (3) geographic divisions with neighbourhood characteristics. The public use sample selected for this thesis is the one-in-a-hundred sample identifying the state as the common geographic unit. The one-in-a-hundred state public use samples for the states in the New England Region are contained on two one-in-a-hundred sample tape files. File 101 tape number T14187 contains the samples for Massachusetts and Rhode Island, and File 102 tape number T14188 contains the samples for Connecticut, Vermont, New Hampshire and Maine.

Each file consists of a large number of 120-character records. For each state sample on the files there are two types of 120-character records: (1) household unit records and (2) person records. Each household unit record is followed by a variable number of person records. There is one person record for each member of the household. If a household unit is vacant, there are no following person records. A household unit record is designated by a 1 in the 120th character location, and a person record by a 3 in the 120th character location.

The data items on the two types of records in the public use sample are presented in Tables 5.3.1 and 5.3.2. Table 5.3.1 shows the data items collected on the basic household unit and identifies the character location of data items prefixed with an H, within the household unit record. Table 5.3.2 shows the data items collected on each person within a household and identifies their character location, prefixed with a P, within the person records.

Table 5.3.1

Items on 1970 Public Use Sample Household Unit Record

Character Location	15-percent sample items	Character Location	15-percent sample items
H1	File number	H66	b̸
H2 H3 H4 H5 H6	Serial number for household or persons in group quarters	H67	b̸
H7 H8 H9 H10 H11	Geographic identifiers (unique for CG, State, and NC samples)	H68	b̸
H12 H13	Persons in household or group quarters	H69	b̸
	HOUSING UNIT CHARACTERISTICS		HOUSEHOLD SUMMARY
H14 H15	Units at this address	H70	Primary family type recode
H16	Vacancy status	H71	Race of head
H17	Duration of vacancy	H72 H73	Age of head
H18	Telephone available	H74	Spanish surname head
H19	Access to unit	H75	Spanish head recode
H20	Complete kitchen facilities	H76	Persons under 18 years of age in household
H21	Rooms	H77	Related children under 18 years of age
H22	Hot and cold piped water	H78	Related children under 6 years of age
H23	Flush toilet	H79	Persons 65 years of age or over
H24	Bathtub or shower	H80	Persons over 60 years of age or over
H25	Plumbing facilities	H81	Roomers, boarders, or lodgers
H26	Basement	H82	Nonrelatives
H27	Tenure	H83	Other relatives
H28	Commercial use	H84	Persons per room
H29 H30	Value	H85 H86 H87	Total family income or income of primary individual
H31 H32 H33	Contract rent	H88	Rent/income
H34 H35 H36	Gross rent	H89	Value/income ratio
H37 H38 H39 H40	Electricity	H90	Year moved in (head)
H41 H42 H43	Gas	H91	Peurto Rican stock (head)
H44 H45 H46	Water	H92	Spanish mother tongue
H47 H48 H49 H50	Fuel	H93 H94	.01 subsample number
H51	Heating equipment	H95 H96 H97 H98 H99 H100 H101 H102 H103 H104 H105 H106 H107 H108 H109 H110 H111	15-percent and 5-percent housing allocations
H52	Year structure built	H112 H113 H114	15-percent allocations
H53	Units in structure	H115	b̸
H54	Location of structure	H116	b̸
H55	Sales of Farm products	H117	b̸
H56	Source of water	H118	b̸
H57	Sewage disposal	H119	b̸
H58	Bathrooms	H20	Record type indicator (1)
H59	Air conditioning		
H60	Autobiles		
H61	b̸		
H62	b̸		
H63	b̸		
H64	b̸		
H65	b̸		

Note: b̸ represents a blank character

Table 5.3.2

Items on 1970 Public Use Sample Persons Records

Character Location	15-percent data sample items
P1	Basic relationship
P2	Detailed relationship
P3	Subfamily number
P4	Type of group quarters
P5	
P6	Sex
P7	Colour or race
P8	Spanish surname
P9	
P10	Age
P11	
P12	Quarter of birth
P13	Marital status
P14	Place of birth recode
P15	State or country of birth
P16	
P17	Highest grade attended
P18	
P19	Finished grade
P20	Children ever born
P21	
P22	Hours worked
P23	Year last worked
P24	
P25	Current industry
P26	
P27	
P28	Current occupation
P29	
P30	Class of worker
P31	Employment status recode
P32	Working 5 years ago
P33	In armed forces 5 years ago
P34	In college 5 years ago
P35	Worked last year
P36	Weeks worked
P37	
P38	Earnings; Wages, etc.
P39	
P40	
P41	Earnings: Business, etc.
P42	
P43	
P44	Earnings: Farm, etc.
P45	
P46	
P47	Social Security Income
P48	
P49	
P50	Welfare income
P51	
P52	
P53	Other income
P54	
P55	Chief Income Recipient
P56	
P57	Person's Total Income
P58	
P59	Poverty Recode
P60	Poverty cutoff
P61	

Character Location	15-percent data sample items
P62	Persons in family or subfamily
P63	
P64	Subfamily relationship
P65	Type of subfamily
P66	Family Unit membership
P67	Family type recode
P68	Family relationship summary
P69	
P70	Parent's birthplace recode
P71	Parents' country of birth
P72	
P73	Mother tongue
P74	
P75	Year moved in
P76	Residence 5 years ago
P77	State of residence
P78	5 years ago
P79	Public or private school
P80	Veteran
P81	Veteran of Vietnam
P82	Veteran of Korean War
P83	Veteran of World War II
P84	Veteran of World War I
P85	Veteran of other time
P86	Place of work: SMSA
P87	b̸
P88	b̸
P89	Place of work: State
P90	Means of transportation to work
P91	Puerto Rican stock
P92	Spanish mother tongue
P93	Spanish recode
P94	Number of times married
P95	Quarter of first marriage
P96	Metro. residence 5 years ago
P97	b̸
P98	b̸
P99	b̸
P100	b̸
P101	b̸
P102	b̸
P103	b̸
P104	
P105	
P106	
P107	
P108	15 percent and 5 percent
P109	population allocations
P110	
P111	
P112	
P113	
P114	
P115	
P116	
P117	15-percent allocations
P118	
P119	
P120	Record type indicator (3)

Note: b̸ represents a blank character

Table 5.3.3

Items from the 1970 Public Use Sample for the Reformatted Record

New Character Location	Data Items	Previous Location
R1	File Number	H1
R2-6	Serial Number for Household	H2-6
R7-8	Serial Number for person in household	*
R9-11	Current Occupation	P27-29
R12-13	State Identifier	H7-8
R14	Urban/Rural Residence	H9
R15	Metropolitan/Non-metropolitan Residence	H10
R16	Central City/non-central city residence	H11
R17	Sex	P6
R18	Colour or race	P7
R19	Spanish-American Recode	P93
R20-22	Age	P9-11
R23	Marital Status	P13
R24-25	Highest Grade Attended	P17-18
R26-28	Current Industry	P24-26
R29	Class of worker	P30
R30	Employment Status Recode	P31
R31	Worked Last Year	P35
R32	Weeks Worked	P36
R33	Hours Worked	P22
R34	Chief Income Recipient	P55
R35-37	Person's Total Income	P56-58
R38-40	Earnings, Wages, etc.	P37-39
R41-43	Earnings, Business, etc.	P40-42
R44-46	Earnings, Farm, etc.	P43-45
R47-49	Social Security Income	P46-48
R50-52	Welfare Income	P49-51
R53-55	Other Income	P52-54
R56	Poverty Recode	P59
R57	Place of Work	P86
R58	Means of Transportation to Work	P90
R59	Basic Relationship	P1
R60	Detailed Relationship	P2
R61-62	Children Ever Born	P20-21

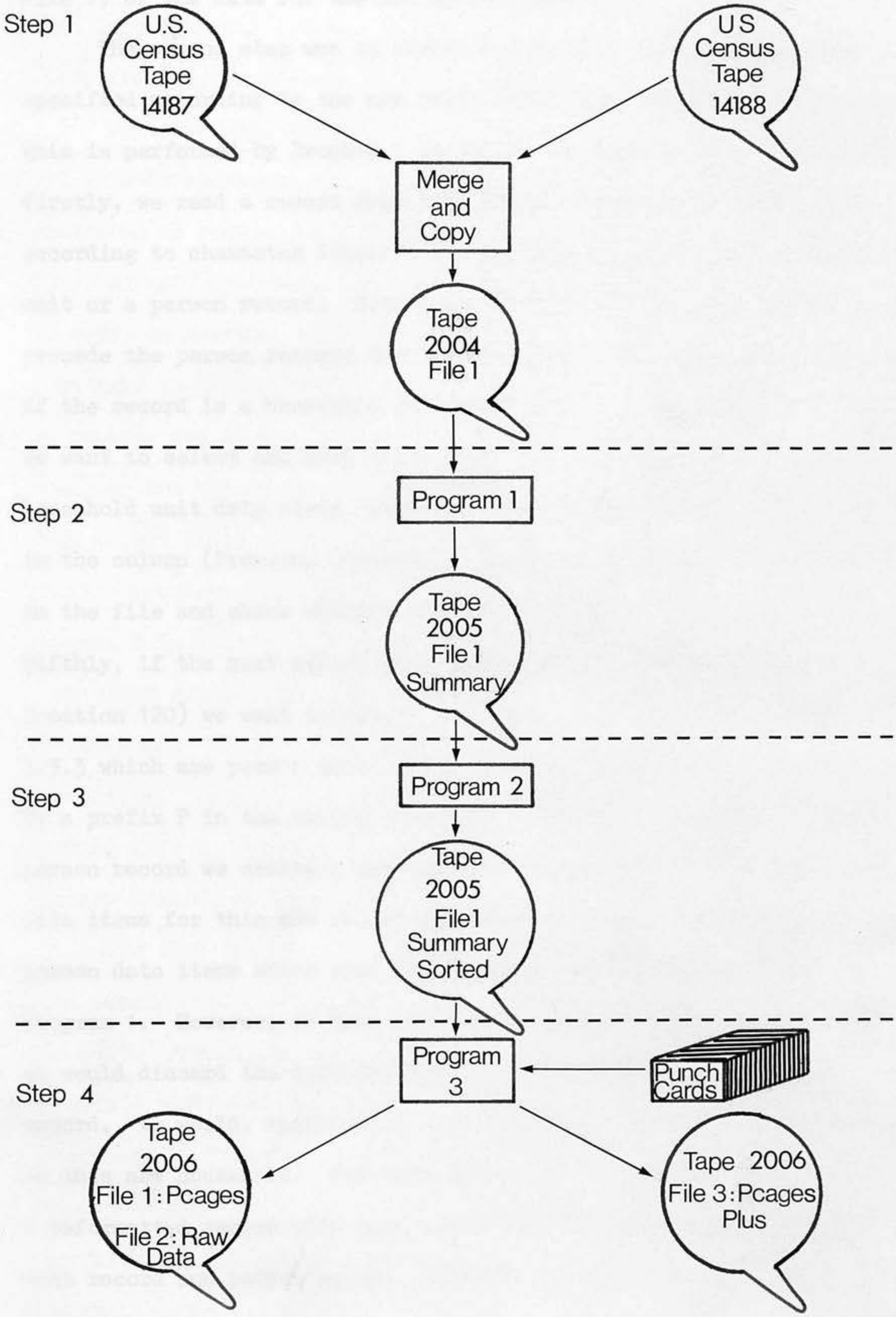
* New code to identify each person within a household

For the purposes of this thesis the data items provided on the household unit and person records are too many in number and too varied in scope to be of relevance and use in this thesis. Therefore, the task in the first stage was to determine which data items would be necessary and possibly useful for an analysis of labour market segmentation. Since we are concerned with individuals, their characteristics, and factors affecting them, we found that more of the relevant data items were on the person records than the household unit record. However, several bits of information from the household unit common to all the persons within the household were required. Table 5.3.3 shows the list of data items selected to form the basic reformatted record for the empirical work in this thesis. The column (Previous Location) indicates by the prefix H or P whether the data item was selected from the household unit record or the person record. The column (New Character Location) designates by the prefix R and a number the order and the location of the data item within the basic reformatted record. After selecting the data items we felt were necessary in order to conduct our analysis, we needed to transform the original data into a form that would enable us to carry out the analysis.

Therefore, in the second stage, it was necessary to develop a computer data processing procedure that would transfer the data items from the appropriate public use sample record to the new reformatted character record. The computer data processing procedure is outlined in diagrammatic form in Diagram 5.3.1. The first step, Step 1, of Diagram 5.3.1, was to ensure an unaltered data set existed. To do so we copied the public use sample Files 101 and 102 onto another tape, namely, tape number 2004. File 101 tape number 14187 and File 102 tape number 14188 became the permanent archive of the public use sample

Diagram 5.3.1

Computer Data Processing Procedure



data which would never be used in the processing and could be referred to if the newly created file were damaged. Tape 2004 is the working file, File 1, of the data for the New England Region.

The second step was to create a file that contained the data specified according to the new basic reformatted records in Table 5.3.3. This is performed by Program 1 in Step 2 on Diagram 5.3.1. In Program 1, firstly, we read a record from tape 2004. Secondly, we determine according to character location 120 whether the record is a household unit or a person record. Note that a household unit record will always precede the person records for the members of the household. Thirdly, if the record is a household unit record (a 1 in character location 120), we want to select and keep those data items on Table 5.3.3 which are household unit data items, that is, those items denoted by a prefix H in the column (Previous Location). Fourthly, we read the next record on the file and check whether it is an household or a person record. Fifthly, if the next record is a person record (a 3 in character location 120) we want to select and keep those data items on Table 5.3.3 which are person data items, that is, those items designated by a prefix P in the column (Previous Location). Finally, for each person record we create a new reformatted record for this person. The data items for this new record are those household unit data items and person data items which were selected and kept in stages 3 and 5 of Program 1. However, if this next record were an household unit record, we would discard the selected data from the previous household unit record. We would, then, select and keep those specified data items on this new household. For each person in a household unit we create a reformatted record with data items from its accompanying household unit record and person record. Program 1 continues until a new reformatted record is created for each person record designated in the original public use sample on tape 2004. These new records are written onto Tape 2005 and located in the working file 1; Summary, of Tape 2005. There are 118,198 records in the file Summary.

Since we want occupation data, the next step, Step 3 on Diagram 5.3.1 is to aggregate the persons from Tape 2005 file 1 Summary according to occupation. There are up to 442 occupational classifications. Thus, we want to create a new file that can accommodate 442 records. However, in creating the new file, we do not want all 118,198 records from Tape 2005 file Summary because some of these records are records on persons not in the labour force, for example, either a person too young or too old to be active in the labour market. Therefore, we will only select those records for persons who are over 14 years of age and in the experienced civilian labour force or in the labour reserve.[4)] Program 2 in Step 3 in Diagram 5.3.1 created the new file. In Program 2 we first removed the records not to be used in creating the file on occupations. Then, we sorted and organised the remaining records by occupation. Program 2 wrote the sorted file onto Tape 2005 in the working file 2, Summary Sorted. There are 66,340 records in file 2, Summary Sorted, on Tape 2005.

Now that the records are in order according to occupation, we must create a file of up to 442 occupational classifications from the 66,340 records on Tape 2005 file 2, Summary Sorted. Program 3 in Step 4 performed three basic operations to create the file on occupations. Firstly, it sorted and summarised the data from the records on each occupation. This data summarized on each occupation is written on Tape 2006 in working file 2, Raw Data. Secondly, Program 3 computed percentages and means for the data items on each

---

4) For the definitions of these terms see Section 5.4 of this Chapter.

occupation. These percentages and means, or summary statistics, were written on Tape 2006 in working file 1, P-cages. Thirdly Program 3 read some additional data items on occupations from punch cards.[5)] Program 3 created another file on Tape 2005 in working file 3 P-cages Plus, that consisted of the data on Tape 2006 file 1 P-cages and the data on punch cards. Program 3 created three separate occupational files. There are 428 occupations in each file; the same 428 occupations.

We must point out that although the U.S. Department of Commerce (1972a) recognises 442 occupational classifications, we are only using 428 in the analyses. This is because in the New England Region 1970 Public Use Sample fourteen of the occupational classifications did not appear. This is not to say these fourteen occupational classifications do not exist in the New England Region, but that in the one-in-a-hundred sample these fourteen did not appear. A detailed list of the 442 occupational classifications appears in Appendix A. The fourteen occupational classifications not included in the New England Region 1970 Public Use Sample are indicated by an asterisk next to their occupational codes in Appendix A.

The third stage involved the introduction of checks or precautions at each step of the data processing procedure to ensure an accurate handling and transferral of the data. The type of precaution taken was to hand-check a sample of the data at each step. For example, to check the transferral of data from Step 1 to Step 2 of Diagram 5.3.1 we followed the procedure that is presented in the next two paragraphs. Table 5.3.4 is one page of the listing of the data is it is used in Step 1 and appears on Tape 2004 File 1. The data columns are enumerated

---

5) For a description of all the data items on occupations see Chapter 7 Sections 7.2 and 7.3.

Table 5.3.4

Basic Public Use Sample Records - Printout of File 1

```
    5    10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95   100  105  110  115  120
100001250000501760001500000003111000000030030020003000052000013100000000001027312220000120607831184000000000000000001400001
0-9990030272012516013672071530100105060999999999999991060445053501002140130001001111200001110020000000000100000000000003
1-9991030261012514102836996960601116999999999999999999900994450035010100001310002222225004911100200000000000070000000000003
2-9990030033612500313870000007033326998998998998998099844500350102000013799022222250049111003000000000000000000000000003
2-9991030001612521313870000007033326998998998998998099844500350102000010799422222250049111003000000000000000000000000003
340990030241512514113471891530110100999999999999999990099944500350107214013000000111120000111002000000000000000000000003
100002250000202760003000002311115157010000510000320053060013000000000000367731000110010084381114600000000000000000000001
0-9991030771212514101850000307611116999999999999015008002472202350400814011100022222250049111002000000000400000000000003
340991030060151251410081927391260110306099999999999999106072200350407814011100022222250049111002000000000000000000000003
100003250000301760005000000311000000300300200030000520600130000000000106731000110000085785113400000000000000000000000001
0-9990030670043510113837889354602216999999999914999060107562903350200899315000011111150049111002000000000010000000000003
1-9991030613012510102850000007611116999999999999999999909996290035020183531500002222225004911100200000000000000000000003
2-9990030202512510113816874530611100000199999000899999900106290035020463501500001111115004911100200000000000000000002003
100004250000402760005000000331100000030030020003000045306001300000000003031313320000010806801111000000000000000000014000001
0-9991030310412514103810673723612202080999999999999999108063704350300000010100022222250049111002000000000001000000000003
2-9991030072612504013870000007033326998998998998998099863700350302000010100122222250049111002000000000000000000000003
2-9991030051612502013870000007033326998998998998998099863700350302000010100122222250049111002000000000000000000000003
2-9990030020612521313870000007033326998998998998998099863700350302000010799422222250049111003000000000000000000000003
100005250000102760001000002311085085100100100010000531000132000000000005059310000000000999680116900000000000000402044000001
0-9991030590212514103850000007622216999999999999999991999018006637088140101000222222500491110020000000000140000000000003
100006251110100760006000000111000000030030020003000050200113300000000004052310000000001207851127000000000000000000000000001
0-9990030523146014113808884150601105120999999999999991120720006637088603015000001161150049111002000000000000000000000003
100007251110200760005000000209000000030030020003000050100123000000000001083310002100000757641107000000000400000000001400001
0-9990030831013716113870000007611116999999999026999048107572202350200000014000011111150049111002000000000000000300000003
1-9991030812010818104860000007611116999999999999999999909997220035020100001400002222225004911100200000000000000000000003
100008251110030522003000304211000000030030020003000740060113400000000000901420000000699878822200000000000000000000000001
100009250110200760005000000208000000030030020003000120060123200000000003054310000000000597641135000000000000000000000001
0-9991030254321251410247539372010220504599999999999999104562302350400835314000022222230000111002000000000000000000000003
2-9991030025151251410017628310010220101399999999999999001362300350404000014000022222230000111002000000000000000000002003
100010251110300760004000000205000000030030020003000130061113100000000001031311110000010767301125000000000000000000000001
0-9990030311012514113478970120102205075999999999999900010766300335010000001010000111103000111002000000000000000000000003
1-9991030231012514101828389260602216999999999999999999909996300035010100006010002222225004911100200000000000000000000003
2-9990030013612521313870000007033326998998998998998099863000350102000010799422222250049111003000000000000000000000003
100011251110100760003000000204000000030030020003000100100110100000000005021310000000000527401157000000000000000000000001
0-9991030211512514100574673720101105052999999999999999105261800663708000010200022222230000111000000000000000000000000003
100012250000201760004000002311080122009009100002900520600130000000000010603100001000012418411145000000000000000000000001
0-9990030600012514213477189033101105120999999999999903112472502350200635014000011111110003111002000000000000000000000003
1-9991030600012511102850000007611116999999999999999999909997250035020183531400002222225004911100200000000000000000000003
```

by 5's. A 1 in column 120 denotes a household unit record and a 3 a person record. The sample data presented in Table 5.3.4 represents the data that were the input for Step 2 of the data processing procedure.

In Step 2, new reformatted records for each person record were created on Tape 2005 File 1 Summary. An example of the listing of new records is shown in Table 5.3.5 and the data columns enumerated by 5's. Referring to Table 5.3.3 we see that the first six character of a record in Table 5.3.5 correspond to the first six characters of the respective household record in Table 5.3.4. Again, referring to Table 5.3.3 we find that the next two characters of a record in Table 5.3.5 indicate the identification of the person record which follows the respective household unit record in Table 5.3.4. The remaining data items on the new record as shown in Table 5.3.5 were checked against the data items in the previous character location in the original records in Table 5.3.4 as designated by Table 5.3.3.

For example, record 10000902 denoted by the arrow on Table 5.3.5 corresponds to the second person record within household unit 100009 denoted by the arrow on Table 5.3.4. The data in character locations R20-22 of record 10000902 in Table 5.3.5 is the age of the person. The data in these characters are 026 and enclosed by the rectangle on Table 5.3.5. Referring to Table 5.3.3 we see that these data items were obtained from character locations P9-11 of the person record. These items are enclosed by a rectangle on Table 5.3.4 and are 026, the same as in the new record. The character location R14 in the new record 10000902 is the data on urban/rural residence. For record 10000902 this is a 1 and circled on Table 5.3.5. From Table 5.3.3 we see that the corresponding data source is character location

Table 5.3.5

Basic Reformatted Record - Printout of Summary

```
     5    10   15   20   25   30   35   40   45   50   55   60
100001 1153250000010270162070105610600609999999999999999994200-13
100001 26962500010102601469906168099999999999999999999994591-02
100001 30002500000100360000070268099899899899899899984592-13
100001 40002500010100062100070268099899899899899899984592-13
100001 5153250000010245141890100409999999999999999999994203413
100002 10002500010107721400076168002499999999999915008759 0-01
100002 2391250001010605149272603810600609999999999999997593400
100003 19352500000106701078846168107599999999999014999060659 0-13
100003 20002500010106101000076168099999999999999999996591-02
100003 34532500000102051068706008001000199999900899999996592-13
100004 13722500010103141406706028108008099999999999999996590-03
100004 20002500010100760400070268099899899899899899986592-13
100004 30002500010100560200070268099899899899899899986592-13
100004 40002500000100262100070268099899899899899899986592-13
100005 10002500010105921400076168199999999999999999990590-03
100006 14152511100105211488806058112012099999999999999997590-13
100007 10002511100108301600076168107599999999999026999048759 0-13
100007 20002511110108101800076168099999999999999999997591-04
100009 13722511110105421453901054104504599999999999999996300-02
100009 2310251111010265146280101100130139999999999999996302-00
100010 10122511100103101489701054107607599999999999900006310-13
100010 29262511110102301483806168099999999999999999996591-01
100010 30002511100100162100070268099899899899899899986592-13
100011 13722511110102151446701055105205299999999999999996300-00
100012 19032500000106001471831054112412099999999999900037130-13
100012 20002500010106001100076168099999999999999999997591-02
100013 17142500000106901440906018105103599999990159999996590-13
100013 20002500010106801200076168099999999999999999996591-02
100013 33722500010104551440901033002502599999999999999996032-00
100014 14702500000103801460701057114014099999999999999997100-13
100014 29522500010103501485731000000200299999999999999997161-03
100014 30002500010101260900070268099899899899899899987592-13
100014 40002500010101160800070268099899899899899899987592-13
100014 50002500010100660300070268099899899899899899987592-13
100015 16522500100103401422801055110009699999999999900036200-13
100015 23102500110103301462801040001701799999999999999996201-03
100015 30002500110101060600070268099899899899899899986592-13
100015 40002500110100860500070268099899899899899899986592-13
100015 50002500110100660300070268099899899899899899986592-13
```

H9 of the household unit record 10009. This data item is a 1 and is circled on Table 5.3.4 in the household record two lines above the arrow.

Steps 3 and 4 of Diagram 5.3.1 was checked in a different manner. The step itself involved reorganising 66,390 records into summary statistics for 428 occupations. The number of records involved makes it next to impossible to attempt any such summary by hand. Therefore, to check that the program operated as desired, we performed a test run on five records from Tape 2005 file 1 Summary. We put the five records through the program, created the summary data on the occupations, and then checked the computations on the five records by hand.

An example of a listing of the data on Tape 2005 File 2 Summary Sorted has not been presented because it resembles the listing shown in Table 5.3.5. Examples of the printouts of the data on Tape 2006 File 2 Raw Data and File 1 P-cages are presented in Tables 5.3.6 and 5.3.7, respectively. These are presented to illustrate the forms that the data took before it was possible to use the data in the analysis. Table 5.3.6 presents one page of computer printout of the summary information on the occupations, and Table 5.3.7 presents one page of computer printout of the summary statistics on the occupational records. Each occupation record in Table 5.3.6 contains 7 rows with 9 counters in the first 6 rows and 8 in the last row. The number to the left on the first row is the occupation identification code.[6)] We treat the counters as being labelled consecutively from the beginning of row one to the end of row seven. Each occupation has 62 items. The first 56 items are counters on descriptive characteristics and the last are counters on subtotals in the occupation. These last counters are used as the denominators in the computation of the summary

---

6) The occupational identification codes are listed in Appendix A.

Table 5.3.6

Basic Data on Occupations - Printout of Row Data

1	349	151	489	9	2	4	21296	8160	84
	361	57	82	380	79	266	115	190	100
	134	346	61	23	1	0	14	113	20
	47	59	32	8	1	148	57	0	365
	78	56	3	43	44	375	13	31	389
	0	379	51258	415	69	0	0	0	40
	3	10	500	151	424	430	462	433	
2	48	5	51	0	2	0	2341	885	3
	43	4	6	38	10	26	11	12	16
	16	32	5	7	9	0	1	6	0
	0	0	0	0	0	37	0	0	27
	5	21	0	3	7	39	2	3	41
	0	48	7028	34	17	3	0	0	5
	0	1	53	5	44	44	49	46	
3	90	32	120	1	1	4	3485	2054	11
	73	5	44	94	17	65	27	22	36
	38	80	9	9	0	0	1	48	5
	5	18	17	0	0	24	4	0	111
	10	1	3	17	21	79	6	8	85
	0	96	9433	116	2	0	0	0	3
	1	6	122	32	96	98	117	99	
4	54	10	63	1	0	0	2186	1105	3
	50	1	13	39	19	35	16	17	22
	22	55	6	0	0	0	1	28	3
	4	14	6	0	0	1	7	0	55
	9	0	1	0	6	56	1	7	53
	0	56	7469	62	1	0	0	0	0
	0	1	64	10	61	61	62	61	
5	12	3	13	0	2	0	520	258	2
	10	2	3	10	4	7	0	3	4
	7	14	0	0	0	0	0	8	0
	1	2	1	0	0	3	0	0	15
	0	0	1	0	1	13	0	0	14
	0	13	1696	14	0	0	0	0	0
	0	0	15	3	14	14	14	14	
6	72	0	70	0	2	0	2710	1237	0
	65	3	4	46	24	29	5	1	42
	23	65	0	0	0	0	0	66	0
	0	0	1	0	0	1	4	0	68
	4	0	4	2	1	68	0	6	60
	0	71	9592	71	0	1	0	0	0
	0	0	72	0	66	65	71	66	
10	22	0	20	1	1	1	864	408	0
	16	2	4	13	9	14	5	3	6
	10	16	2	1	0	0	1	16	0
	0	0	1	0	0	3	1	0	20
	2	0	0	2	0	19	0	1	18
	0	22	3929	21	3	0	0	0	1
	0	0	22	0	19	19	21	19	
11	129	2	127	2	2	1	5601	2200	1
	112	7	12	93	24	71	27	41	19
	46	87	15	3	1	0	60	8	12
	0	0	1	0	0	32	17	0	73
	51	7	3	10	7	107	1	8	104
	0	119	15578	120	9	0	0	0	7
	1	0	131	2	106	105	124	113	
12	273	8	273	4	4	2	10518	4950	3
	233	7	41	195	68	150	33	57	95
	103	245	11	5	0	0	8	191	38
	4	0	7	0	0	24	9	0	254
	21	6	8	9	23	244	1	7	253
	0	261	38293	276	7	2	1	0	7
	0	2	281	8	255	261	276	261	

Table 5.3.7

56 Summary Statistics on Occupations - Printout of P-cages

1	69,800	30,200	97,800	1,800	0,400	0,800	42,592	16,320	0,556	72,200
	11,400	16,400	76,000	15,800	53,200	23,000	0,448	0,236	0,316	0,805
	0,142	0,053	0,200	0,0	2,800	22,600	4,000	9,400	11,800	6,400
	1,600	0,200	29,600	11,400	0,0	73,000	15,600	11,200	0,600	0,093
	0,095	0,812	0,030	0,072	0,898	0,0	75,800	132,516	83,000	13,800
	0,0	0,0	0,0	8,000	0,600	2,000				
2	90,566	9,434	96,226	0,0	3,774	0,0	44,170	16,698	0,600	81,132
	7,547	11,321	71,698	18,868	49,057	20,755	0,273	0,364	0,364	0,727
	0,114	0,159	16,981	0,0	1,887	11,321	0,0	0,0	0,0	0,0
	0,0	0,0	69,811	0,0	0,0	50,943	9,434	39,623	0,0	0,061
	0,143	0,796	0,043	0,065	0,891	0,0	90,566	132,604	64,151	32,075
	5,660	0,0	0,0	9,434	0,0	1,887				
3	73,770	26,230	98,361	0,820	0,820	3,279	28,566	16,836	0,344	59,836
	4,098	36,066	77,049	13,934	53,279	22,131	0,229	0,375	0,396	0,816
	0,092	0,092	0,0	0,0	0,820	39,344	4,098	4,098	14,754	13,934
	0,0	0,0	19,672	3,279	0,0	90,984	8,197	0,820	2,459	0,145
	0,179	0,675	0,061	0,081	0,859	0,0	78,689	77,320	95,082	1,639
	0,0	0,0	0,0	2,459	0,820	4,918				
4	84,375	15,625	98,438	1,563	0,0	0,0	34,156	17,266	0,300	78,125
	1,563	20,313	60,938	29,688	54,688	25,000	0,279	0,361	0,361	0,902
	0,098	0,0	0,0	0,0	1,563	43,750	4,688	6,250	21,875	9,375
	0,0	0,0	1,563	10,938	0,0	85,938	14,063	0,0	1,563	0,0
	0,097	0,903	0,016	0,115	0,869	0,0	87,500	116,703	96,875	1,563
	0,0	0,0	0,0	0,0	0,0	1,563				
5	80,000	20,000	86,667	0,0	13,333	0,0	34,667	17,200	0,667	66,667
	13,333	20,000	66,667	26,667	46,667	0,0	0,214	0,286	0,500	1,000
	0,0	0,0	0,0	0,0	0,0	53,333	0,0	6,667	13,333	6,667
	0,0	0,0	20,000	0,0	0,0	100,000	0,0	0,0	6,667	0,0
	0,071	0,929	0,0	0,0	1,000	0,0	86,667	113,067	93,333	0,0
	0,0	0,0	0,0	0,0	0,0	0,0				
6	100,000	0,0	97,222	0,0	2,778	0,0	37,639	17,181	0,0	90,278
	4,167	5,556	63,889	33,333	40,278	6,944	0,015	0,636	0,348	1,000
	0,0	0,0	0,0	0,0	0,0	91,667	0,0	0,0	0,0	1,389
	0,0	0,0	1,389	5,556	0,0	94,444	5,556	0,0	5,556	0,028
	0,014	0,958	0,0	0,091	0,909	0,0	98,611	133,222	98,611	0,0
	1,389	0,0	0,0	0,0	0,0	0,0				
10	100,000	0,0	90,909	4,545	4,545	4,545	39,273	18,545	0,0	72,727
	9,091	18,182	59,091	40,909	63,636	22,727	0,158	0,316	0,526	0,842
	0,105	0,053	0,0	0,0	4,545	72,727	0,0	0,0	0,0	4,545
	0,0	0,0	13,636	4,545	0,0	90,909	9,091	0,0	0,0	0,095
	0,0	0,905	0,0	0,053	0,947	0,0	100,000	178,591	95,455	13,636
	0,0	0,0	0,0	4,545	0,0	0,0				
11	98,473	1,527	96,947	1,527	1,527	0,763	42,756	16,794	0,500	85,496
	5,344	9,160	70,992	18,321	54,198	20,611	0,387	0,179	0,434	0,829
	0,143	0,029	0,763	0,0	45,802	6,107	9,160	0,0	0,0	0,763
	0,0	0,0	24,427	12,977	0,0	55,725	38,931	5,344	2,290	0,061
	0,056	0,863	0,009	0,071	0,920	0,0	93,840	118,916	91,603	6,870
	0,0	0,0	0,0	5,344	0,763	0,0				
12	97,153	2,847	97,153	1,423	1,423	0,712	37,431	17,616	0,375	82,918
	2,491	14,591	69,395	24,199	53,381	11,744	0,224	0,373	0,404	0,939
	0,042	0,019	0,0	0,0	2,847	67,972	13,523	1,423	0,0	2,491
	0,0	0,0	8,541	3,203	0,0	90,391	7,473	2,135	2,847	0,033
	0,083	0,884	0,004	0,027	0,969	0,0	92,883	136,274	98,221	2,491
	0,712	0,356	0,0	2,491	0,0	0,712				

Table 5.3.8

60 Summary Statistics on Occupations - Printout of P-cages Plus

1	69,800	30,200	97,800	1,800	0,400	0,800	42,592	16.320	55,629	72,200
	11,400	16,400	76,000	15,800	53,200	23,000	44,811	23.585	31,604	80,465
	14,186	5,349	0,200	0,0	2,800	22,600	4,000	9.400	11,800	6,400
	1,600	0,200	29,600	11,400	0,0	73,000	15,600	11.200	0,600	9,307
	9,524	81,169	3,002	7,159	89,838	0,0	75,800	102.516	83,000	13,800
	0,0	0,0	0,0	8,000	0,600	2,000	1,000	0.0	0,0	0,0
2	90,566	9,434	96,226	0,0	3,774	0,0	44,170	16.698	60,000	81,132
	7,547	11,321	71,698	18,868	49,057	20,755	27,273	36,364	36,364	72,727
	11,364	15,909	16,981	0,0	1,887	11,321	0,0	0,0	0,0	0,0
	0,0	0,0	69,811	0,0	0,0	50,943	9,434	39.623	0,0	6,122
	14,286	79,592	4,348	6,522	89,130	0,0	90,566	132.604	64,151	32,075
	5,660	0,0	0,0	9,434	0,0	1,887	1,000	0,0	0,0	0,0
3	73,770	26,230	98,361	0,820	0,820	3,279	28,566	16.836	34,375	59,836
	4,098	36,066	77,049	13,934	53,279	22,131	22,917	37.500	39,583	81,033
	9,184	9,184	0,0	0,0	0,820	39,344	4,098	4,098	14,754	13,934
	0,0	0,0	19,672	3,279	0,0	90,984	8,197	8.820	2,459	14.530
	17,949	67,521	6,061	8,081	85,859	0,0	78,689	77.320	95,082	1,039
	0,0	0,0	0,0	2,459	0,820	4,918	1,000	0,0	0,0	0,0
4	84,375	15,625	98,438	1,563	0,0	0,0	34,156	17.266	30,000	78,125
	1,563	20,313	60,938	29,688	54,688	25,000	27,869	36,066	36,066	90,164
	9,836	0,0	0,0	0,0	1,563	43,750	4,688	6,250	21,875	9,375
	0,0	0,0	1,563	10,938	0,0	85,938	14,063	0,0	1,563	0,0
	9,677	90,323	1,639	11,475	86,885	0,0	87,500	116,703	96,875	1,563
	0,0	0,0	0,0	0,0	0,0	1,563	1,000	0,0	0,0	0,0
5	80,000	20,000	86,667	0,0	13,333	0,0	34,667	17.200	66,667	66,667
	13,333	20,000	66,667	26,667	46,667	0,0	21,429	28,571	50,000	100,000
	0,0	0,0	0,0	0,0	0,0	53,333	0,0	6,667	13,333	6,667
	0,0	0,0	20,000	0,0	0,0	100,000	0,0	0,0	6,667	0,0
	7,143	92,857	0,0	0,0	100,000	0,0	86,667	113,067	93,333	0,0
	0,0	0,0	0,0	0,0	0,0	0,0	1,000	0,0	0,0	0,0
6	100,000	0,0	97,222	0,0	2,778	0,0	37,639	17.181	0,0	90,278
	4,167	5,556	63,889	33,333	40,278	6,944	1,515	63,636	34,848	100,000
	0,0	0,0	0,0	0,0	0,0	91,667	0,0	0,0	0,0	1,389
	0,0	0,0	1,389	5,556	0,0	94,444	5,556	0,0	5,556	2,817
	1,408	95,775	0,0	9,091	90,909	0,0	98,611	133,222	98,611	0,0
	1,389	0,0	0,0	0,0	0,0	0,0	1,000	0,0	0,0	0,0
10	100,000	0,0	90,909	4,545	4,545	4,545	39,273	18,545	0,0	72,727
	9,091	18,182	59,091	40,909	63,636	22,727	15,789	31,579	52,632	84,211
	10,526	5,263	0,0	0,0	4,545	72,727	0,0	0,0	0,0	4,545
	0,0	0,0	13,636	4,545	0,0	90,909	9,091	0,0	0,0	9,524
	0,0	90,476	0,0	5,263	94,737	0,0	100,000	178,591	95,455	13,636
	0,0	0,0	0,0	4,545	0,0	0,0	1,000	0.0	0,0	0,0
11	98,473	1,527	96,947	1,527	1,527	0,763	42,756	16.794	50,000	85.496
	5,344	9,160	70,992	18,321	54,198	20,611	38,679	17.925	43,396	82,857
	14,286	2,857	0,763	0,0	45,802	6,107	9,160	0,0	0,0	0,763
	0,0	0,0	24,427	12,977	0,0	55,725	38,931	5.344	2,290	8,065
	5,645	86,290	0,885	7,080	92,035	0,0	90,840	118.916	91,603	6,870
	0,0	0,0	0,0	5,344	0,763	0,0	1.000	0,0	0,0	0,0
12	97,153	2,847	97,153	1,423	1,423	0,712	37.431	17,616	37,500	82,918
	2,491	14,591	69,395	24,199	53,381	11,744	22,353	37,255	40,392	93,870
	4,215	1,916	0,0	0,0	2,847	67,972	13,523	1,423	0,0	2,491
	0,0	0,0	8,541	3,203	0,0	90,391	7,473	2.135	2.847	3,261
	8,333	88,406	0,383	2,682	96,935	0,0	92,083	136.274	98,221	2,491
	0,712	0,356	0,0	2,491	0,0	0,712	1,000	0.0	0,0	0,0

statistics on the first 56 items. Each occupation in Table 5.3.7 contains six rows. There are ten statistics in the first five rows and six in the sixth row. The statistics are treated as though they were labelled consecutively from the beginning of row one to the end of row six. These 56 data items are the summary statistics compiled from the New England Regionone-in-a-hundred public use sample. The final group of data items for our analysis are contained on Tape 2005 File 3 P-cages Plus. An example of one page of computer printout of Tape 2005 File 3 P-cages Plus represents the form in which the data will be used in the analysis and is shown in Table 5.3.8. The organisation of the occupational records is the same as that for the files on Tape 2006, except in the File 3 P-cages Plus there are four additional data items. The file P-cages Plus contains six rows in which there are ten summary statistics.

In summary, we developed a systematic procedure to select the relevant data items, to transform the original data into a form appropriate for our analysis, and to check that the data was processed accurately. As a result of these procedures, we created a tape file (Tape 2005 File 3 P-cages Plus) of 428 occupational classifications. Each occupational classification on the file has 60 summary statistics. Thus, from the 1970 New England Public Use Sample we have created a data bank of 60 summary statistics observed from each of 428 occupational classifications that is appropriate for our analysis of labour market segmentation.

5.4 Definitions

Table 5.4.1 provides a list of the terms and the subject concepts recognised in the 1970 Census. The terms and the subject concepts included are those items that appear in the list of data items in Tables 5.3.1, 5.3.2, and 5.3.3 and require further explanation for use in the empirical work. The number in parentheses after the term refers to the character location of the data in the new reformatted record listed in Table 5.3.3. The list is designed to provide a reference of information about the nature of the data items. The terms and concepts in Table 5.4.1 are derived from the responses to one or more of the census questions, and in most cases are specific responses.

Table 5.4.1

Concepts Recognised in the 1970 Census
(U.S.Department of Commerce, 1972b, pp. 129-172)

1. Current Occupation (R9-11) - ascertained for persons 14 years of age and over in the experienced labour force or in the labour reserve. Employed persons were to report the occupation at which they worked the most hours during the reference week. The experienced unemployed and persons in the labour reserve were to report their last occupation. The detailed occupation classification scheme identifies 442 occupations.

   (a) Employed (R9-11, R28-29, R30) - civilians 14 years and over who during the reference week were either "at work" - who did any work for pay or profit or worked without pay for 15 hours or more on a family farm or business; or "with a job but not at work" - were temporarily absent due to illness, vacation, etc.

   (b) Unemployed (R9-11, R28-29, R30) - civilians 14 years and over who were neither "at work" nor "with a job but not at work" but were looking for work.

   (c) Experienced Unemployed (R9-11,R26-28, R30) - those unemployed who indicate they have worked at some time in the past.

   (d) Experienced Civilian Labour Force (R9-11, R28-29, R30) - includes all those in categories 1a, 1b, and 1c.

   (e) Not in labour force (R9-11, R28-29, R30) - persons 14 and over not included in 1a, 1b, 1c, or 1d.

   (f) Labour Reserve (R9-11, R26-28, R30) - persons not in the labour force but employed within the last ten years.

2. Unemployment Rate (R30) - represents the number of unemployed as a percent of the experienced civilian labour force.

3. Urban/Rural Residence (R14) - towns and incorporated, or unincorporated places of 2,500 or more inhabitants; all other places are rural.

4. Urbanised Area (R15, R16, R57) - an area containing a city of 50,000 or more inhabitants plus the surrounding closely settled incorporated and unincorporated areas.

5. Metropolitan/Nonmetropolitan Residence (R15) - metropolitan refers to persons residing in a Standard Metropolitan Statistical Area (SMSA); areas with one city with 50,000 or more inhabitants, with two cities with contiguous boundaries and a combined population of at least 50,000, or with a population density of at least 100 persons per square mile;

Table 5.4.1 con'd.

nonmetropolitan are all other areas.

6. Central City/Non-Central City Residence (R16) - acity with a population over 50,000 and within an urbanised area.

7. Race (R18) - refers to the division of the population into White, Negro or Black, and several other racial categories.

8. Spanish-American Recode (R19) - persons of Spanish language.

9. Marital Status (R23) - persons were asked whether they were "now married", "widowed", "divorced", "separated", or "never married".

10. Highest Grade Attended (R24-25) - for persons 5 years and over, the highest grade or year of regular school they ever attended up to 6 or more years of college; those in school indicated the year they were completing.

11. Current Industry (R26-28) - ascertained for persons 14 years of age and over in the experienced civilian labour force. Employed persons were to report the job at which they worked the most hours during the reference week. The experienced unemployed and persons in the labour reserve were to report the job that they last held. The industry classification scheme in this thesis has 13 major industry groups which were identified by the Census Bureau.

12. Class of Worker (R29) - the class of work is indicated by one of the following categories - (1) private wage and salary workers, (2) government workers, (3) self-employed workers, (4) unpaid family workers.

13. Worked Last Year (R31) - all persons 14 years and over who worked during the previous calendar year.

14. Weeks Worked (R32) - the number of weeks worked during the previous calendar year.

15. Hours Worked (R33) - the number of hours worked during the reference week.

16. Chief Income Recipient (R34) - for a household unit, the family member with the largest income; for a group quarter, every person or unrelated person over 14.

17. Total Income (R35-37) - the sum of the dollar amounts of money respondents reported receiving as wages or salary income, net farm income, self-employment income, and other income.

(a) Wages and Salary (R38-40) - money income received as

Table 5.4.1 con'd.

wages, salary, commissions, bonuses, or tips from all jobs before deductions of taxes, dues, etc. Excludes pay in kind.

(b) Self-Employed earnings (R41-43, R44-46) - money income received as profits or fees from businesses, professional practice, partnership, or farm.

(c) Social Security Income (R47-49) - includes government payments to retired persons, to dependents of deceased insured workers, or to disabled workers; excludes medicine.

(d) Welfare Income (R50-52) - includes amounts received from Federal, State, and local public programs such as aid for dependent children, old-age assistance, general assistance, and aid to the blind or totally disabled.

(e) Other Income (R53-55) - includes interest, dividends, veteran payments, retirement pensions, and other regular payments such as unemployment insurance benefits, workmen's compensation, etc. Excludes receipts from sale of personal property, capital gains on payments in kind.

18. Poverty Recode (R56) - families and unrelated individuals are classified as being below the poverty level if the total family income is less than three times the cost of the U.S. Department of Agriculture "economy food plan". This plan takes into account family size, number of children, area of residence, as well as income.

19. Place of Work (R57) - ascertained for persons 14 and over who reported working at some time during the reference week. They were asked where they worked. Persons who worked at more than one job are to report place of work for the job at which they worked the greatest number of hours; persons who travelled in their work or worked in more than one place are to report where they began work or where they worked the most hours.

(a) Urban Fringe (R57) - the incorporated and the unincorporated areas of more than 2,500 inhabitants surrounding the central city of an urbanised area.

(b) Central Business District (R57) - the downtown retail trade area of a city; defined only in cities with a population over 100,000. These areas are characterised by a high concentration of retail business offices, theatres, hotels, and service businesses.

(c) Ring of an SMSA (R57) - all of the SMSA not part of the central city.

Table 5.4.1 con'd.

20.	Means of Transportation to Work (R58) - ascertained for persons 14 years and over who reported working during the reference week. Principal mode of travel or type of conveyance used to get to their place of work on the last day they worked.
21.	Children Ever Born (R61-62) - the total number of live births for all women aged 14 or over, regardless of marital status.

CHAPTER 6

Techniques of Analysis

6.1 Classification

The techniques of analysis used in this thesis lie within the broad subject area of classification. Classification, in the broad sense of the term, allocates observations either into initially undefined groups or to predefined groups so that observations are (according to some criterion) "close" to one another. In classification our data are either the entire population or a sample of observations from the population. Each observation possesses a set of measurable characteristics, and it is these that provide the basis for the analysis.

We shall use two techniques within the broad subject of classification. The first technique, namely, classification analysis, refers to the process of subdividing the observations into distinct groups. The other technique which we shall use is discriminant analysis. Discriminant analysis provides a rule for allocating new observations to one of several predefined groups, and, as an extension, for discriminating among groups.

Classification, then, is a broad subject area of techniques which enables us to attach labels to groups of observations. So, to a degree, classification provides us with statistical techniques for clarifying and organising information on groups of observations. An important practical consequence of the process of clarification and organisation is that an investigation of the data can communicate information on the data. Thus, the most important purpose of adopting classification techniques is for organising information in a revealing way. These techniques also facilitate the development of inductive generalisations and the generation of possible hypotheses about the information which had not been evident.

6.2 Classification Analysis

In the specific technique of classification analysis, our intentions are to perform three functions: firstly, develop a criterion for delineating groups; secondly, decide how many groups; and thirdly, allocate the observations into the groups. Classification problems arise when a sample or population is suspected to have a certain number of groups, but the origins of the groups are unknown. Our purpose is to determine what the subdivisions are.

According to Kendall and Stuart (1966, p. 336), "the problem of classification, as we define the word, is one of determining from empirical evidence whether individuals 'group' or 'cluster'." Kendall and Stuart (1966, p. 336) point out two ways of looking at the classification problem.

(1) Given, as usual, a p x n vector of observations, let us consider the n sample points in the p-dimensional Euclidean space determined by the p variables. If these points, to some acceptable definition, fall into clearly distinguishable groups, we may say that the n individuals may be <u>classified</u> into those groups. Their 'nearness' is to be considered as a function of the variate values which they bear.

(2) In the alternative p-space embedded in an n-space the variables are represented by vectors. There is some interest in how far these vectors cluster, as we have seen in canonical analysis. In this case we are concerned with the extent to which the variables cluster, not the individuals.

In the first way of looking at classification problems, we accept the variables and are primarily interested in classifying or grouping the observations. In the second we are concerned with the importance of the variables, that is, whether they are all necessary and which are more important. In this thesis we are concerned in the first instance with using classification analysis to identify the groups, and in the second instance with determining the importance of each variable within any group.

The issue to resolve in classification analysis is to define what is meant by a group. There are many views on what constitutes a group. All of the views rely upon some kind of measure for determining the distance between observations. We can use such a measure to decide upon the closeness of any two observations. As Cormack (1971, p. 329) pointed out, two basic ideas are involved: internal cohesion and external isolation. Internal cohesion stresses a high correlation among members of a group; whereas external isolation emphasises the degree of discontinuity between groups. The consequence is that we must specify some measures for determining internal cohesion and external isolation, and thus, for identifying the groups.

In defining groups, most classification techniques optimise some criterion over the sample of observations or the population. The techniques rely upon algorithms as a method of optimising some criterion function. Cormack (1971, p. 330) specified three types of algorithms for obtaining classes:

1. agglomerative - series of fusions of the observations into groups;
2. divisive - partition of a complete population set successively into smaller groups;
3. clustering - successive re-allocation of observations between the possible sets of some initial population.

The three methods generate groups at many levels of significance, and the number of groups used are obtained by choosing the desirable level of significance. Clustering methods in particular, may specify the desirable properties of the observations for each group. According to Cormack (1971, p. 330) the importance of the specification of workable algorithms is that "algorithms can be carried through and an answer obtained, whereas other better defined methods can perhaps not be implemented."

6.3 Hierarchical Classification

For the purposes of this thesis, we want to take into consideration the similarities of the observations on occupations with respect to many properties. The method of generating previously undefined groups used in this thesis is hierarchical classification. It is a general classification process by which N observations with p properties are joined progressively to form groups. After selecting the properties of the observations to be used in the analysis, the next step in the procedure of hierarchical classification requires the specification of a measure of similarity or dissimilarity between every pair of observations in the population. The two most similar observations are grouped, and then the similarity of this new observation group is computed with the other observations. The procedure will continue to classify all the observations until they have all been joined into one large group.

The method of hierarchical classification adopted in this thesis was presented by Ward (1963). The method establishes each possible number of groups, g, g-1, ... 1, according to a measure of similarity and in a manner that minimises the loss of information associated with each grouping. In hierarchical classification we must specify an appropriate measure of similarity implying some notion of distance between observations. Ward used the Euclidean distance as the measure of similarity or dissimilarity between two observations.[1] The Euclidean distance is a distance measure most appropriately used with variables that are all on a similar continuous scale. The data in this thesis are of this type. In addition, he specified an algorithm that would present the loss of information associated with each step in the process of fusing observations into groups. The algorithm used in the Ward method seeks to minimise the error sum-of-squares. So, at each step, the union of every possible pair of observations is considered, and the resulting grouping is that which produces the minimum increase in the error sum-of-squares.

The grouping procedure starts with a population set, N, consisting of n_n one-element subsets or groups. The subsets from the population N are designated by S_g, where the subscript g identifies the subset. In presenting the technique of hierarchical classification we shall adopt the approach used by Wishart (1969). In general the error sum-of-squares, E_g, for the group S_g is

$$E_g = \sum_{i=1}^{n_g} \sum_{j=1}^{p} (X_{ijg} - \overline{X}_{jg})^2 \; ; \qquad (6.3.1)$$

where X_{ijg} is the value of the j-th variable for the i-th observation in group S_g, containing n_g observations, and $\overline{X}_{jg}$ is the mean value of the j-th variable in S_g. When expanded, E_g becomes

---

1) Euclidean distance = $\sum_{v=1}^{p} w_v (X_{iv} - X_{jv})^2$ where w_v is used to normalise the variables; v refers to the variable; X refers to the observations with the subscripts i and j, denoting specific observations.

$$E_g = \sum_{i=1}^{n_g} \sum_{j=1}^{p} X^2_{ijg} - n_g U^2_g \; ; \qquad (6.3.2)$$

where U^2_g is the inner product $U'_g U_g$, and $U'_g = (\overline{X}_{1g}, \overline{X}_{2g}, \ldots \overline{X}_{pg})$ is the vector of means for S_g.

The desirability of the groupings is reflected in the value of the objective function E, that is the sum of the error sum-of-squares for each of the g groups.

$$E = \sum_{i=1}^{g} E_g \, . \qquad (6.3.3)$$

If we are trying to join groups S_a and S_b to create a new group S_c, we can write the increase in E as

$$I_{ab} = E_c - E_a - E_b \qquad (6.3.4)$$

where E_c is the error sum-of-squares for the union set $S_c = S_a \cup S_b$. Using (6.3.1) we can rewrite (6.3.4) as

$$I_{ab} = \sum_{i=1}^{n_c} \sum_{j=1}^{p} X^2_{ijc} - n_c U^2_c - \sum_{i=1}^{n_c} \sum_{j=1}^{p} X^2_{ija} + n_a U^2_a -$$

$$\sum_{i=1}^{n_b} \sum_{j=1}^{p} X^2_{ijb} + n_b U^2_b \; . \qquad (6.3.5)$$

In (6.3.5) the sums of the squares X^2_{ij} cancel, thus

$$I_{ab} = n_a U^2_a + n_b U^2_b - n_c U^2_c \; . \qquad (6.3.6)$$

From matrix algebra we can obtain an expression for $U^2{}_c$. Since

$$n_c{}^2 \, U^2{}_c = (n_a U_a + n_b U_b)^2,$$

$$= n_a{}^2 U^2{}_a + n^2{}_b \, U^2{}_b + 2 \, n_a n_b \, U'{}_a \, U_b$$

$$= n_a{}^2 \, U^2{}_a + n^2{}_b \, U^2{}_b + n_a n_b \, (U^2{}_a + U^2{}_b - (U_a - U_b)^2), \qquad (6.3.7)$$

we can reduce $U_c{}^2$ to the following:

$$U_c{}^2 = \frac{n_a}{n_c} U^2{}_a + \frac{n_b}{n_c} U^2{}_b - \frac{n_a n_b}{n^2{}_c} (U_a - U_b)^2 . \qquad (6.3.8)$$

Now we can substitute $U_c{}^2$ into equation (6.3.6) which becomes

$$I_{ab} = (n_a n_b / n_c) \, (U_a - U_b)^2 , \qquad (6.3.9)$$

but $(U_a - U_b)^2$ is the distance between the means of the groups S_a and S_b, in other words the Euclidean distance, $d^2{}_{ab}$, between the means. Thus, we have

$$I_{ab} = (n_a n_b / n_c) \, d^2{}_{ab} \qquad (6.3.10)$$

and any joining of groups will occur when I_{ab} is a minimum. This is a direct consequence of the Euclidean distance being a minimum.

The first step of the algorithm used by Ward is to calculate the Euclidean distance, d^2 above, between every possible pair of observations. The second step is to determine the increase in the error sum-of-squares, from (6.3.10) above, which results from the fusion of any two observations. The third step is to join those two observations for which d^2 is a minimum and thus I is a minimum. The process continues in this manner until all the observations are reclassified from n groups to one group.

The E.R.C.C. computer program used for the hierarchical classification relies upon the method due to Ward (1963).[2)] The method represents the observations and the grouping process in a dendogram. A dendogram represents a geometric interpretation which shows the value of the Euclidean distance, the similarity, for each join and the error sum-of-squares, the loss of information, resulting at each join. An analysis of a dendogram discloses the similarities and dissimilarities among the subsets of each group at any point in the process.

In addition, the method generates an F-ratio and a T-ratio. The F-ratio is the ratio of the standard deviation of a variable within a group to the standard deviation of the variable over the entire population. This ratio provides a measure of internal cohesion, or put differently, of the "compactness" (closeness to one another) of the within group variances. To define the T-ratio one takes the difference between the mean value of a variable within a group and its mean over the population and divides this by the mean of the variable over the population. The T-ratio provides a measure of external isolation of the group. The greater is the value of the T-ratio, the greater is the difference between the means, and hence, the more isolated or "distinct" (distance from the others) is the group.

---

2) See Appendix B.

6.4 Example of Hierarchical Classification

To illustrate the procedure of hierarchical classification, we present an example given by Ward (1963, p. 241-3). In the simple example, he considered the problem of grouping five individuals on the basis of ratings they had given to one (unidentified) object. The individuals and their respective object ratings are presented in Table 6.4.1. The algorithm used is the same as that described in section 6.3. The method, then, was one that would minimise the error sum-of-squares.

Figure 6.4.1 illustrates the dendogram that summarises the hierarchical classification. As we can see from Figure 6.4.1, initially there were five distinct groups. At the next stage of the classification, persons 1 and 3 were joined and each of the other persons constituted three separate groups. From the dendogram we can see that the classification successively fused persons and groups until all five persons constituted one group.

The values in the three rows below the dendogram are particularly worth noting in this example of hierarchical classification. The values of the entries in row E, the error sum-of-squares, represent the loss of information associated with each successive stage in the classification. Further understanding of the assignment of individuals to the groups is facilitated by an examination of the $\triangle 1$ entries and the $\triangle 2$ entries. The $\triangle 1$ entries represent the change in the error term at each classification, and the $\triangle 2$ entries indicate the acceleration of error with each classification. From the E entries we can conclude that as the number of groups approaches one group, the loss of information increases. Furthermore, looking at $\triangle 1$ and $\triangle 2$ entries, we can conclude that the change in the error and the acceleration of error becomes progressively larger in the process of forming one group. The

Table 6.4.1

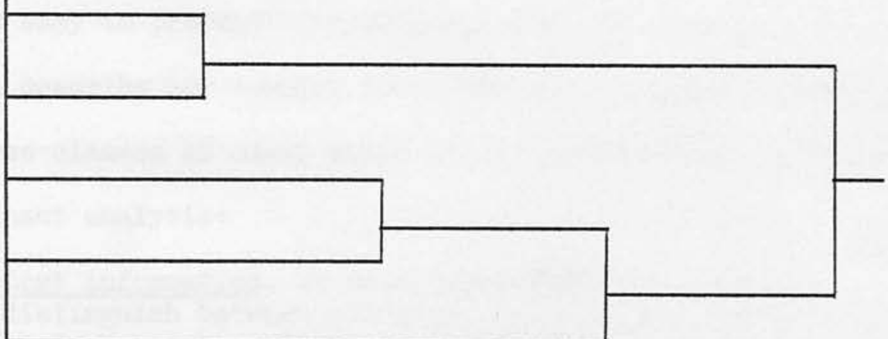
Numerical Example: Individuals and the Object Ratings

Person	1	2	3	4	5
Object Rating	1	7	2	9	12

Table 6.4.2

Dendogram of Results of Hierarchical Classification

Person	Number of Groups								
	5		4		3		2		1
1									
3									
2									
4									
5									
E	.00		.50		2.50		13.17		86.80
△1		.50		2.00		10.67		73.63	
△2			1.50		8.67		62.96		



acceleration of error at the classification into two groups increased markedly over that of three groups.

6.5 Discriminant Analysis

In the technique of discriminant analysis we are faced with the problem of assigning observations to one of several known and defined samples or populations. The purpose of discriminant analysis, then, is to provide a criterion for allocating an observation to one of the groups. Discriminant analysis requires the provision of an initial set of groups possessing a set of measurable characteristics. The technique then uses these characteristics to estimate equations from the initial set of groups. Thus, discriminant analysis serves to classify new observations to the group whose characteristics are most similar. In addition, discriminant analysis provides a measure of the probability that an observation is a member of the group to which it is assigned. With this we can investigate the membership and the characteristics of the groups, as well as the quality of the groupings.

It is easy to present circumstances that may require discriminant analysis to describe how samples were separated. Kendall (1966, p. 167) defined three classes of cases which may be investigated by the technique of discriminant analysis:

(1) Lost information. It would have been very easy to distinguish between different races at the time when their bones were in the living body, but if all that we have to go on today are archaeological remains , we may need indirect observations or surviving material to provide a method of discriminating between alternative attributions of newly discovered data.

(2) Diagnosis. Nature may provide us with information about the hidden presence of disease, by surgical necessity or post-mortem. But we really require to diagnose from external symptoms the presence or absence unambiguously to be determined are not, in general,available. In fact,

it is often our object to detect the condition without being driven to employ such methods.

(3) Prediction. It may be possible to differentiate two conditions without error when they occur, but we may want to discriminate before they occur. For example, if movements up or down in unemployment rates were found to be correlated with certain economic movements which lead them in time, such as capital investment, it would be desirable in advance so that remedial measures can be put in hand.

Originally, the technique of discriminant analysis was developed to predict the accuracy of assigning an observation to one of several known groups on the basis of a set of measurable characteristics. The problem centred upon a class of alternative statistical decision hypotheses. Only one decision could be accepted, and the others were to be rejected. In the case of two populations the investigators had to test the balance between incorrect and correct decisions on choosing one hypothesis rather than the other. Even though discriminant analysis may not be completely successful in classifying every new observation, it does reduce the possibility of an incorrect classification. The accuracy of assignment depends in part upon the set of characteristics on which discriminant analysis is based.

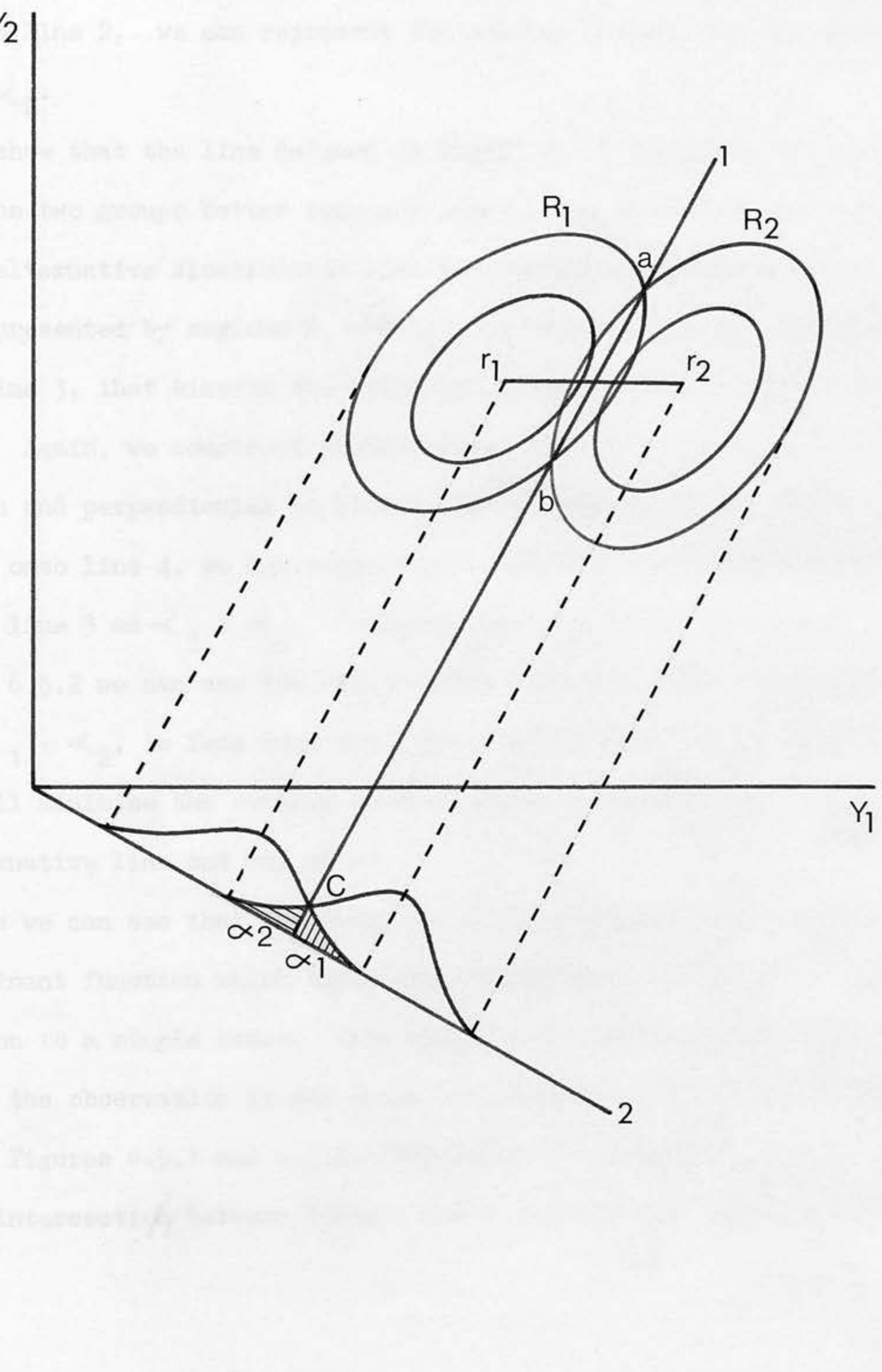
With discriminant analysis we can represent the criterion for classification or for the identification of mutually exclusive groups as a linear combination of the set of characteristics on the observations in the known groups. In the case of two groups in two dimensions the linear function may be regarded as a line in an two-dimensional Euclidean space in which each point in the space is associated with each observation. The line serves as the optimal partition between groups of points in the space. For more than two groups and large numbers of characteristics we can more appropriately describe discriminant analysis as a technique for fitting hyperplanes through an Euclidean space to partition it into distinct sections. We may use the significance of the variables in the linear function to identify the important dimensions in the discriminating space.

To be more specific, we can use the technique of discriminant analysis to consider a set of N observations possessing p characteristics. The N observations are arranged in two groups, G_1 and G_2. The groups, G_1 and G_2, contain N_1 and N_2 members, respectively, such that $N_1 + N_2 = N$. Each member of the two groups has an accompanying vector whose elements, x_1, x_2, ... x_p, correspond to the values for the p characteristics. Generally, we assume that the populations of G_1 and G_2 are distributed multivariate normally. However, the component characteristics on the populations may not be normally distributed. The technique of discriminant analysis does offer a solution to this problem. According to Cooley and Lohnes (1962, p. 116), "since by the Central Limit Theorem the linear functions of variates are more likely to be normal than are the component variates, multiple-discriminant scores may satisfy the important assumption of a multivariate normal distribution better than the original test scores."

The N_1 and N_2 members of groups G_1 and G_2 may be associated with points in a p-dimensional Euclidean space. The geometric interpretation of discriminant analysis in two dimensions, Y_1 and Y_2, is represented in Figure 6.5.1. The region of each ellipse, R_1 and R_2 in Figure 6.5.1 represents the locus of points of equal density for each group, G_1 and G_2, respectively. For example, the outer ellipse for R_1 may be defined as the region within which 90 percent of group G_1 lies, and the inner ellipse for R_1 may be defined as the region within which 75 percent of the group lies. The points, r_1 and r_2, represent the centroids of the ellipses.

Figure 6.5.1

Projections of Two Groups on the Best Discriminant Lines
(Cooley and Lohnes, 1962, p. 117)



In discriminant analysis we must determine a smooth and defined boundary which minimises the overlap in partitioning regions R_1 and R_2. The two points a and b in Figure 6.5.1 correspond to the points of intersection of the two outer ellipses. We will use these two points of intersection to define a straight line, line 1, partitioning the groups. If we construct a second line, line 2, originating from the origin and perpendicular to line 1, and if we project the points in the space onto line 2, we can represent the overlap between the two groups as $\alpha_1 + \alpha_2$.

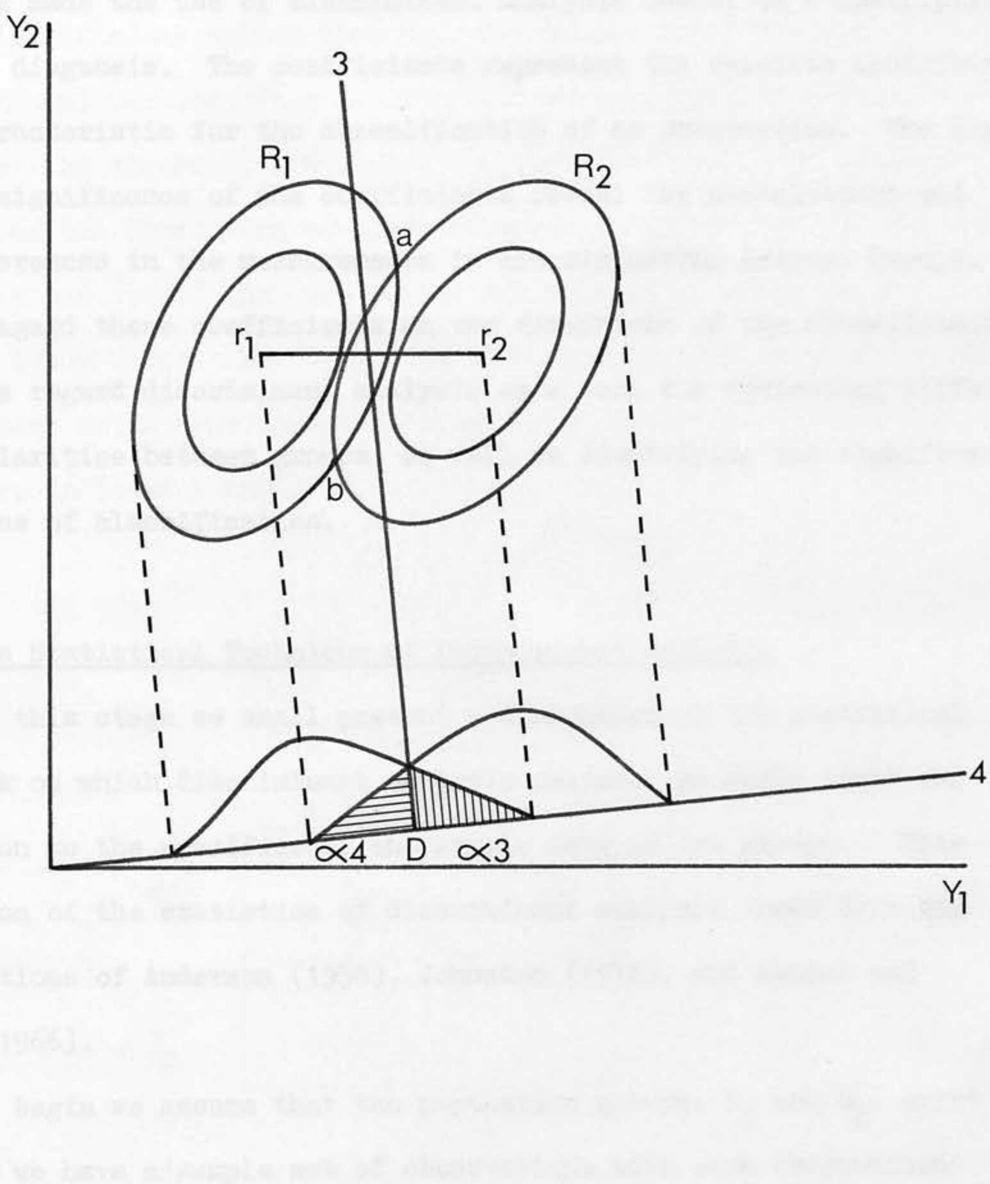
To show that the line defined in Figure 6.5.1 minimises the overlap between the two groups better than any other line, we present in Figure 6.5.2 an alternative discriminant line for partitioning our groups G_1 and G_2 represented by regions R_1 and R_2. In Figure 6.5.2 we construct a line, line 3, that bisects the line connecting the centroids of the ellipses. Again, we construct another line, line 4, originating from the origin and perpendicular to line 3. After projecting the points in the space onto line 4, we can represent the overlap resulting from the partition line 3 as $\alpha_3 + \alpha_4$. Comparing the partitions in Figures 6.5.1 and 6.5.2 we can see that the overlap resulting from the discriminant line 1, $\alpha_1 + \alpha_2$, is less than that from line 3, $\alpha_3 + \alpha_4$. Therefore, line 1 will minimise the overlap between these two groups better than this alternative line and any other.

Thus we can see that the technique of discriminant analysis generates a discriminant function which transforms the p characteristics of an observation to a single score. This score is the transformation of the point for the observation in the space to a position along line 2 and line 4 in Figures 6.5.1 and 6.5.2. The point c in Figure 6.5.1, the point of intersection between lines 1 and 2, divides the line 2 into two

Figure 6.5.2

Alternative Discriminant Line

(Rulon et al., 1967, p. 302)



sections. The left side represents the probable membership of G_1 and the right of G_2. Point d in Figure 6.5.2 similarly indicates the probable membership of G_1 and G_2 resulting from the alternative discriminant line.

In addition, with discriminant analysis we can determine the coefficients of the variables in the linear functions. It is this which has made the use of discriminant analysis useful as a descriptive tool for diagnosis. The coefficients represent the relative contribution of a characteristic for the classification of an observation. The size and the significance of the coefficients reveal the associations and the differences in the measurements in discriminating between groups. We can regard these coefficients as the dimensions of the classifications. Thus, we regard discriminant analysis as a tool for indicating differences and similarities between groups, as well as identifying the significant dimensions of classification.

6.6 The Statistical Technique of Discriminant Analysis

In this stage we shall present a discussion of the statistical framework on which discriminant analysis relies. We shall limit our discussion to the specifics of the simple case of two groups. This discussion of the statistics of discriminant analysis draws from the presentations of Anderson (1958), Johnston (1972), and Kendal and Stuart (1966).

To begin we assume that two population groups, G_1 and G_2, exist and that we have a sample set of observations with each observation possessing p measurable characteristics, x_1, x_2, ... x_p. On the basis of the p characteristics we want to divide the p-dimensional space into two regions, R_1 and R_2, as for Figure 6.5.1. We then wish to set up a rule that when an observation possesses a certain set of values of the p characteristics, it may be associated with a point in region R_1, and therefore we assign it to population G_1. If it has other values, the point associated with it falls in R_2, and we classify it as coming from G_2.

With this classification rule, we can make two kinds of error. One source of error results when an observation is actually from G_1 and is classified as coming from G_2. The second error results if it is from G_2 and is classified as from G_1. We have presented the costs of correct and incorrect classification in Table 6.6.1. We denote the cost of the first kind of classification error as c (2/1), box 2 of Table 6.6.1, and the cost of the second kind of error as c (1/2), box 3 of Table 6.6.1. Both costs of misclassification are greater than zero. There is no cost involved in making a correct classification, the zeroes are in boxes 1 and 4.

Table 6.6.1

Costs of Classification

		Decision: G_1	Decision: G_2
Group	G_1	1 0	2 c(2/1)
	G_2	3 c(1/2)	4 0

In discriminant analysis we intend to minimise the expected costs of misclassification. Therefore, we must consider some way of defining what is meant by minimising the costs of misclassification. One way of doing so is to begin by making the assumption that we have the a priori probabilities of the two groups. The probability that an observation comes from G_1 is p_1 and from G_2 is p_2. We assume that the properties of the groups can be specified by density functions. We let the density of G_1 be f_1 (x), and that of G_2 be f_2 (x).

Since we have specified a region R_1 of classification as from G_1 and a region R_2 as from G_2, we can write the probability of misclassifying an observation from G_1 as

$$P\ (2/1,\ R) = \int_{R_2} f_1\ (x)\ dx\ , \qquad (6.6.1)$$

where $dx = dx_1, \ldots .\ dx_p$. Similarly, we can write the probability of misclassifying an observation from G_2 as

$$P\ (1/2,\ R) = \int_{R_1} f_2\ (x)\ dx\ . \qquad (6.6.2)$$

The probability of drawing an observation from G_1 is p_1, and the probability of drawing an observation from G_1 and misclassifying it is p_1 P (2/1, R), that is, the probability associated with box 2 in Table 6.6.1. Likewise, the probability of drawing an observation from G_2 is p_2, and the probability of drawing an observation from G_1 and misclassifying it is p_2 P (1/2, R), the probability associated with box 4 in Table 6.6.1.

The expected cost of misclassification is the sum of the product of each cost of misclassification multiplied by the probability of occurrence and is given by

$$c\ (2/1)\ p_1 \int_{R_2} f_1\ (x)\ dx + c\ (1/2)\ p_2 \int_{R_1} f_2\ (x)\ dx. \qquad (6.6.3)$$

Since we can derive

$$\int_{R_1} f_2\,(x)\,dx = 1 - \int_{R_2} f_2\,(x)\,dx = \int f_2\,(x)\,dx - \int_{R_2} f_2\,(x)\,dx, \tag{6.6.4}$$

we can express (6.6.3) as

$$\int_{R_2} \Big[c\,(2/1)\,p_1 f_1\,(x) - c\,(1/2) p_2 f_2\,(x) \Big]\,dx + \Big[c\,(1/2)\,p_2 \Big] \Big[\int f_2\,(x)\,dx \Big]. \tag{6.6.5}$$

The last term of (6.6.5) is a positive constant. Thus, we can minimise (6.6.5) if R_2 is defined as the set of p characteristics, the x's for which

$$c\,(2/1)\,p_1 f_1\,(x) - c\,(1/2)\,p_2 f_2\,(x) < 0\,. \tag{6.6.6.}$$

We then define R_1 as the set of x's for which the expression in (6.6.6) is positive. Now we can state the classification rule as the ratio of the density functions. The rule is

$$\begin{aligned} &R_1 \text{ if } \frac{f_1(x)}{f_2(x)} > \frac{c\,(1/2)\,p_2}{c\,(2/1)\,p_1} \\ &R_2 \text{ if } \frac{f_1(x)}{f_2(x)} < \frac{c\,(1/2)\,p_2}{c\,(2/1)\,p_1}\,. \end{aligned} \tag{6.6.7}$$

In words, we classify an observation into G_1 or G_2 as the ratio of the likelihood of an observation's score exceeds or falls short of the constant on the right hand side of the expressions in (6.6.7).

However, to apply the expressions in (6.6.7) we must know the values of the a priori probabilities, the costs, the density functions, and the values of the parameters in the density functions. To derive a more practical expression for (6.6.7), we shall assume that our groups are multivariate normal with different mean vectors μ_1 and μ_2 but with the same covariance matrix $\sum$. We further assume that we know the form of the density functions. We then can write the ratio of the likelihoods as

$$\frac{f_1(x)}{f_2(x)} = \frac{\exp\left[-\frac{1}{2}(x-\mu_1)\ \Sigma^{-1}\ (x-\mu_1)\right]}{\exp\left[-\frac{1}{2}(x-\mu_2)\ \Sigma^{-1}\ (x-\mu_2)\right]} \qquad (6.6.8)$$

We can identify (6.6.8) to

$$\frac{f_1(x)}{f_2(x)} = \exp \quad x\,\Sigma^{-1}(\mu_1-\mu_2) - \frac{1}{2}(\mu_1+\mu_2)'\,\Sigma^{-1}(\mu_1-\mu_2) \ .$$

(6.6.9)

If we let

$$\delta = \Sigma^{-1}\ (\mu_1-\mu_2) \qquad (6.6.10)$$

and let k denote the right hand side of the expressions in (6.6.7) we can rewrite the classification rule as

$$R_1 \text{ if } x'\delta - \frac{1}{2}(\mu_1+\mu_2)'\delta > \log k$$

$$R_2 \text{ if } x'\delta - \tfrac{1}{2}(\mu_1 + \mu_2)'\delta < \log k \,. \qquad (6.6.11)$$

we refer to $x'\delta$ as the discriminant function.

In general we apply discriminant analysis to populations or groups which are not known, but must be inferred from samples. Thus we must be able to classify when the parameters are only estimates. To perform discriminant analysis in this case we must make estimates for μ_1 and μ_2 for the classification rule (6.6.11).

We now assume that we have the vectors of sample means, $\bar{x}_1$ and $\bar{x}_2$, computed from the sample sets n_1 and n_2 from our groups G_1 and G_2, respectively. These vectors are the likelihood estimates of μ_1 and μ_2. We can compute the deviations from the sample means X_1 from $\bar{x}_1$ and X_2 from $\bar{x}_2$. We then base our likelihood estimate of the common variance matrix Σ on the pooled sums of squares from the two samples and define it as

$$S = \frac{1}{(n_1 + n_2 - 2)} \, (X_1' X_1 + X_2' X_2). \qquad (6.6.12)$$

When we substitute the likelihood estimates into the left hand side of the expressions in (6.6.11) we obtain

$$x'S^{-1} \, (\bar{x}_1 - \bar{x}_2) - \tfrac{1}{2}(\bar{x}_1 + \bar{x}_2)' \, S^{-1} \, (\bar{x}_1 - \bar{x}_2) \quad . \qquad (6.6.13)$$

The discriminant function becomes

$$Y = x' \, S^{-1} \, (\bar{x}_1 - \bar{x}_2) \quad . \qquad (6.6.14)$$

We want to find the discriminant function that will discriminate most successfully between the observations and their given characteristics. This is the function that has the greatest variance between samples relative to the variance within samples. This is to say, Y is the function that maximises

$$\frac{(\bar{Z}_1 - \bar{Z}_2)^2}{\sum_{i=1}^{2} \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2} , \qquad (6.6.15)$$

where $Z = x' S^{-1} (\bar{x}_1 - \bar{x}_2)$ is a scalar.[3] For any vector of observations $\bar{x}$ it has a mean value $\bar{Z}_1$ taken over the sample from G_1 and a mean value $\bar{Z}_2$ taken over the sample from G_2. In other words, this provides us with a discriminant function that will separate two groups as much as possible by making the distance between the groups as great as possible.

The technique of discriminant analysis used in this thesis is contained in the E.R.C.C. BMDØ7M computer program.[4] This program performs a stepwise discriminant analysis; that is, at each step only one variable is entered into the set of discriminating variables according to the criteria set out on the first page of appendix C. Those variables with F-values below a minimum are deleted from the analysis. From the set of discriminating variables, a discriminant function is specified. A discriminant function is specified for each group in the analysis.

The actual discriminant scores for the observations are not calculated. Instead, the square of the Mahanalobis distance of each observation from each group is calculated. An observation is classified with the group to which it is closer. The Mahanalobis distance between

---

3) See Anderson (1958, p. 136).

4) See Appendix C.

two groups on the basis of p characteristics is

$$D^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} S^{-1} (\bar{x}_{i1} - \bar{x}_{i2}) (\bar{x}_{j1} - \bar{x}_{j2}). \quad (6.6.16)$$

This expression can be manipulated to yield a variance ratio similar to that in expression (6.6.15).[5] The significance of the Mahanalobis distance is that is provides a chi-squared measure of the probability that an observation is a member of the group to which it is assigned. It is this that enables us to make statements about the quality of the classifications.

In addition, the program computes two canonical variables, which are linear combinations of the variables used in the analysis. For each observation the program computes the values for the canonical variables. These values are then used to identify a point in a two-dimensional space associated with each observation. The mean location of each group is also depicted on this space. This provides a visual interpretation of the group constellations.

6.7 The Interpretation of Discriminant Analysis with an Example

There are two general techniques of discriminant analysis. Both techniques rely upon the statistical framework presented above. The first technique is the characteristic root method. This method computes discriminant functions from the vectors associated with the latent roots of the covariance matrix of the characteristics. The maximum number of discriminant functions generated by the characteristic root method is the lesser of two numbers; either one less than the number of groups

5) See Rao (1970, p. 297).

or the number of characteristics on the observations. Each observation has a score for each of the discriminant functions. If one of the observation's scores falls in the range of scores associated with one of the groups, the observation is classified into that group.

The other technique, the one used in this thesis, is the likelihood method, which computes a discriminant function associated with each group. According to Gramm (1973, p. 342) these functions are called likelihood discriminant functions because "the score for a group is proportional to the logarithm of the likelihood of belonging to that group." For each observation there is a score computed on each of the discriminant functions. The observation is classified into the group identified with the function giving the observation the highest score. The discriminant function which produces the highest score for the observation may be considered the line in the space to which the observation lies closer. Both methods usually give the same results. The characteristic root method does so with fewer discriminant functions, but with an accompanying loss of specific information on each group.

With the likelihood method we can obtain three types of information on our groups in a discriminant space: firstly, the centres of gravity of each group; secondly, the metric distance between groups; and thirdly, the dimensions of classification defined by the discriminant functions. To analyse the discriminant functions we rely both on the use of the coefficients of the variables in the functions and on the use of the scores of the observations for each function.

We consider that the score for each discriminant function is influenced by the variables in the function in proportion to the size of their coefficients and in the direction of the sign of their coefficients. Variables with small (positive or negative) coefficients do not significantly effect the discriminant function; whereas variables with large coefficients do. A variable with a large positive coefficient tends to be proportional to the score for the function; whilst a variable with a large negative coefficient tends to be inversely proportional to the score. In other words, the variables are uncorrelated, positively correlated, and negatively correlated with the score for a discriminant function. If an observation has a high value for a variable with a large positive coefficient in a particular function, it is likely to be classified in that group. If an observation has a high value for a large negative coefficient in a function, it is less likely to be classified in that group.

With the likelihood discriminant method we obtain one discriminant function for each group. The similarities and dissimilarities among the groups are expressed in the differences between the sizes of the coefficients for the same variable in the different discriminant functions. After computing the discriminant score for an observation on each of the different functions we can classify the observation to that group for which it had the highest discriminant score. We note that in this thesis we will use the Mahanalobis distance to represent the discriminant score, and thus assign an observation to the group associated with the function for which it had the lowest score (closest in distance). We can conclude which observations have more of the characteristic properties measured by the discriminant functions and which ones do not have those properties.
By determining which variables are more significant in the discriminant functions, we can identify those variables that influence the likelihood of being classified into a particular group. To illustrate the technique of discriminant analysis, we present an analysis performed by Gramm (1973). In this analysis Gramm sought to determine those variables which distinguish among the three labour force alternatives available to married women

teachers. The alternatives are: not to work, part-time work, and full-time work. The basis for the analysis relies upon Gramm's model for the labour force decision of married women teachers. In this model Gramm (1973, p. 343) suggests that the decision to supply labour is derived from the household demand for the leisure of the wife, and is a function of the wage of the husband, wage of the wife, household assets, household age (number of years since marriage), and the ages of the children.

Gramm used the likelihood method of discriminant analysis to classify married women teachers into one of the three labour force categories. From the analysis she obtained the three discriminant functions shown in Table 6.7.1. The size and the sign of the coefficients were used to infer the influence of the variables on the labour force decision. We only present a few of the findings to illustrate the manner of interpretations. Gramm found a large coefficient for the husband's wage for non-workers and a small coefficient for it for full-time workers. She interpreted this to mean that as the husband's wage increases, the more likely the women should be classified as a non-worker. The coefficient for the wife's full-time wage is largest for full-time workers as is the coefficient for the wife's part-time wage. For both variables this was interpreted to mean that as the wage offered increases so does the likelihood that the wife will be classified as a full-time worker. The coefficients of household age were interpreted as an indication that with increasing household age a woman is less likely to work full-time and more likely to work part-time. The coefficients for the ages of children were held to indicate that the older are children, the more likely the woman is to work full-time. Gramm concluded that the ages of the fourth and fifth children

Table 6.7.1

Likelihood Discriminant Functions Families with Children
(Gramm, 1973, p. 344)

Variable	Alternative		
	Non-work	Part-time	Full-time
Wage of husband	.163	.101	-.024
Full-time wage of wife	2.124	2.067	2.614
Part-time wage of wife	9.609	9.924	10.109
Household assets	-.008	.002	-.007
Household age	.094	.092	.037
Age of 1st child	.284	.507	.953
Age of 2nd child	-.250	-.254	-.342
Age of 3rd child	.285	.254	.263
Age of 4th child	-.203	-.338	-.479
Age of 5th child	1.562	1.699	1.832
Constant	-32.665	-35.008	-40.992

Table 6.7.2

Classification Matrices: Families with Children
(Gramm, 1973, p.346)

Likelihood Discriminant Functions

Actual	Predicted: Zero	Predicted: Part-time	Predicted: Full-time	Of observations in each group % correctly classified
Zero	32	10	0	76.2
Part-time	32	58	13	56.3
Full-time	7	14	58	73.4
Percent of all observations correctly classified				66.1

Root Discriminant Functions

Actual	Predicted: Zero	Predicted: Part-time	Predicted: Full-time	Of observations in each group % correctly classified
Zero	29	13	0	69.0
Part-time	32	54	17	52.4
Full-time	2	18	59	74.7
Percent of all observations correctly classified				63.4

are more important than the ages of the other children in discriminating among the labour force status.

The discriminant scores for each observation on each discriminant function were used to assign the observations to one of the labour force groups. It is worth pointing out that Gramm obtained the scores for observations by both the likelihood method and the characteristic root method. We have presented in Table 6.7.2 the classification matrices that Gramm obtained by both methods. The elements of the matrices in Table 6.7.2 show how many correct classifications and misclassifications occurred in using both methods of discriminant analysis. Of all the observations, 66.1 percent were classified correctly by the likelihood method and 6.3.4 percent by the characteristic root method. Overall, the two methods of discriminant analysis produced close predictions of the actual labour force status of the married women teachers in the study.

6.8 Summary

Classification techniques are tools for assigning observations to groups, unknown or defined. However, classification techniques do not lead to perfectly mutually exclusive groups. Most classification techniques generate disjoint groups with the minimum possibility of misclassification. Often, an observation can reasonably be a member of one or more groups. The actual membership depends upon the characteristics used in the analysis and the purpose of the classification: the purpose of any particular classification will in large determine the type of groupings resulting.

The most important purpose of classification is to provide a different viewpoint on the data by creating a different framework in which to view the data. The techniques of classification enable us to investigate the associations and the variations of the characteristics within and among groups. The technique of hierarchical classification creates a set of groups whose origins were previously unknown. Discriminant analysis provides a rule for assigning new observations to one of several known groups and for discriminating between groups. More importantly, discriminant analysis provides the means for measuring the suitability of the classification and identifying the dimensions of classification. Thus, we can regard classification techniques as a type of multi-dimensional scaling for studying the relationships within and among groups in a new and revealing way.

CHAPTER 7

Labour Market Segmentation and Classification Analysis

7.1 The Problem of Devising Labour Market Segments

The concept of labour market segmentation suggests an alternative analytical approach in which to view the structure and operation of the labour market. This approach distinguishes a threefold division of the jobs and labour force in the labour market. Each division or labour market segment is characterised by different labour market characteristics and behaviour. Labour market segmentation offers a structural framework of the labour market from which to organise information about the conditions, income, and development of dominant groups of individuals in the earnings process in terms of the influence of economic, and non-market and non-competitive factors. In other words, the concept of labour market segmentation attempts to incorporate many possible factors operating in the labour market that may affect the allocation and pricing of labour.

However, because of the nature of the description of this concept and the limitations of available data, a straightforward and empirical operational view of labour market segmentation is not easily accommodated. The concept of labour market segmentation is abstract and is not necessarily associated with any specific set of data. But, with the various techniques of classification analysis we can organise the available data into a form compatible and consistent with the notion of labour market segmentation. Classification analyses offer the opportunity for generating possible representations of the labour market segments and for clarifying the broad and general notion of labour market segmentation.

With the technique of classification analysis we can use the available data on occupational classifications (described in Chapter 5) to generate these labour market segments. We are concerned with using a broad range of factors in the classification analysis that may distinguish the three labour market segments and determine how sellers of labour cluster. The factors range from the industrial and occupational characteristics of the labour market to the demographic characteristics of the individuals in the labour market. It is these factors that determine the limits and scope for action of the sellers of labour to enter, exit, and earn income in the labour market segments. Classification analysis provides the means with which to use these factors to define the similar groups of jobs and individuals within the labour market structure.

Therefore, with the technique of classification analysis we go beyond the specification of the labour markets according to one or two variables or dimensions. Instead, as we saw in Chapter 4, we represent the labour market segments in terms of a multi-dimensional classification scheme devised from an analysis of the economic, and non-market and non-competitive factors in operation in the labour market. We regard the techniques of classification analysis as statistical tools to develop an empirical view of labour market segmentation. To recap, by using classification analysis we can provide a more thorough and detailed description of the labour market segments. Thus, we regard, respectively, the concept of labour market segmentation and the techniques of classification analyses as the framework and the tools with which to establish a first approximation to a social, institutional, and economic description of the segments within the labour market.

As previously discussed, the data to be used in the classification analysis are the occupational data on 428 occupational classifications that were prepared from the New England Region 1970 Public Use Sample. The data items on these 428 occupational classifications will be used to specify which occupations are grouped into which labour market segments. These occupational classifications constitute the specific occupations recognised by the U.S. Department of Commerce and are based on the description of the duties and activities required on the job. The U.S. Department of Commerce organises the detailed list of occupational classifications into twelve major occupational groups, that, in turn, are organised into four occupational areas. Table 7.1.1 lists the twelve occupational groups and the arrangement of these into four occupational areas.

In a sense, the four occupational areas could be considered one way to define classes of labour. However, as we shall see in the following discussion, the four occupational areas do not provide any information beyond a list of similarly described jobs. In combining the specific occupational classifications in Appendix A into the twelve major occupational groups, any number of specific occupational classifications (ranging from 3 to create the group Farmers and Farm Managers up to 123 to create the Professional, Technical, and Kindred Workers' group) was used to form the twelve groups.[1] The criterion used in the arrangement into the twelve groups was simply to organise and represent similarly described job duties and activities. These twelve groups are then arranged into four occupational areas of similar types of work.

---

1) See Appendix A for the detailed list of specific occupational classifications and the organisation of these into twelve occupational groups.

Table 7.1.1

U.S. Bureau of the Census

Major Occupational Groups and Occupational Areas

(U.S. Department of Commerce, 1972b, p.153)

1970 Major Occupational Groups	Occupational Areas
1. Professional, Technical, and Kindred Workers	
2. Managers and Administrators, except Farm	White Collar
3. Sales Workers	
4. Clerical and Kindred Workers	
5. Craftsmen and Kindred Workers	
6. Operatives, except Transport	
7. Transport Equipment Operatives	
8. Labourers, Except Farm	Blue Collar
9. Farmers and Farm Managers	
10. Farm Labourers and Farm Foremen	Farm
11. Service Workers, except Private Household	
12. Private Household Workers	Service

But, even though the occupational classifications within an occupational area may be characterised by similarly described activities and duties, this does not mean that responsibilities, qualifications, career levels, and rewards are similar. In fact, the specific occupations within an occupational area cover the spectrum of occupational opportunities at all career levels. Some of the specific occupational classifications may be dead-end jobs, whilst others may be steps along a career ladder. For example, the occupational area Blue Collar includes Sales Workers, but there are specific occupational classifications such as stock and bond salesmen that may be more appropriately grouped within the White Collar occupational area. A bond salesmen's responsibilities and pay may be substantially different from that of another occupational classification in the Sales group, such as that of a newsboy. Thus, we find that the specific occupations within an occupational area do not necessarily require the same or similar duties.

Furthermore, these occupational areas are not determined according to the detailed differences in the characteristics of workers and the treatment of the workers in the labour market. In this thesis we are concerned with specifying groups that are compatible with the notion of labour market segmentation. The specification of these groups cannot ignore the different treatment of different types of labour within the labour market. From the example of the Sales group above, we can conclude that four occupational areas do not organise the workers and jobs according to their treatment within the structure of the labour market. In fact, the reason for developing the concept of labour market segmentation was to describe a labour market in which different workers in the group of non-disadvantaged workers, such as the White Collar occupational area, were treated differently. Therefore, we find that the four occupational areas in Table 7.1.1. do not comply with the purposes of this thesis.

In other words, we want to group occupations that show similar preferences for specific types of labour. In addition, we want the groups to reflect similar conditions and treatment of workers. To do so, we must reorganise the information on the workers in each of the 428 specific occupations and the selected characteristics of the occupations in a manner that represents the concept of labour market segmentation. The techniques of classification analysis will be applied to this data to devise objective criteria for classifying occupations into the labour market segments. Consequently, the criteria will provide a stochastic description of those factors that determine an individual's treatment and position in the labour market, or put differently, the factors influencing the allocation and pricing of labour.

Nevertheless, with the general description of labour market segmentation and the limited amount of data available, we can hardly expect the application of the techniques of classification analysis to provide an exact and complete quantitative analysis of the labour market. Yet, the techniques of classification analysis can be used as a narrowly defined and conditioned application for establishing a more precise notion of labour market segmentation. The statistical formulation of the labour market segments depends on the use of many variables. These represent characteristics of the occupations obtained from observations, but are still too few for the techniques of classification analyses to duplicate exactly the segments as described in the concept. The specification of the occupations into labour market segments is only as good as the type and amount of data available. The number and type of variables is limited by the data available, but the number is sufficient to generate a first approximation to a multi-dimensional description of the labour market segments.

7.2 Procedure

From the range of applications of the techniques of classification analysis we shall use two approaches in considering the concept of labour market segmentation. The first approach relies upon the conceptual description of labour market segmentation to approximate a first estimate of the grouping of a sample from the 428 occupations into the three labour market segments. The first estimate is made according to an objective examination of a limited number of data items and a subjective judgement of the types of occupations. Then we apply the analytical method of stepwise discriminant analysis to derive three linear functions that discriminate among groups of occupations on the basis of the characteristics provided for the analysis. The linear functions maximise the ratio of between groups to within group variances. They maximise the separation between groups. The discriminant analysis identifies those characteristics and attributes of individuals and occupations that correspond to a classification into the independent primary, subordinate primary, or secondary segments of the labour market. On the basis of the score from the linear discriminant functions the occupations in the first estimate are reassigned to the labour market segment that its characteristics and attributes most resemble. Discriminant scores are obtained for the occupations previously unassigned and are used to classify these occupations in the same manner.

The second approach is strictly empirical in substance. This approach uses the technique of hierarchical classification that determines occupational groups statistically by measuring the similarity of occupations according to Euclidean distance and by minimising the error sum-of-squares. The technique considers each occupation in turn, computes its similarity with every occupation and group, and then assigns it to the nearest group. Hierarchical classification generates the best possible classification of occupations from 428 groups to 1 group.

For the purposes of this thesis, we shall use hierarchical classification to generate classifications into three groups. The three groups created by hierarchical classification can then be analysed in terms of the correspondence to the three labour market segments. The final step in this second approach is the application of stepwise discriminant analysis to the set of three groups of occupations resulting from the hierarchical classification. In this approach, discriminant analysis serves to determine whether there are any occupations in a marginal or transitional zone and, if so, whether they are appropriately assigned. In a sense, the application of discriminant analysis to this set of three groups optimises the groups because it involves the more powerful criterion of a chi-square measure for determining the final assignment of individuals.

Thus, the first approach provides a statistical formulation of a first approximation of the conceptual description of labour market segmentation; whereas, the second approach parallels the first as a simple empirical presentation of the groupings of the occupations. On the one hand, in the first approach we restrict the domain of possible groupings of occupations in order to accommodate a fit according to the concept of labour market segmentation. On the other hand, in the second approach we allow hierarchical classification to generate groups empirically from the data. Both the conceptual approach and the empirical approach are used to determine the characteristics of the labour market segments and the quality of the classifications. The parallel approaches provide a systematic approach to develop the labour market segments, to clarify the concept of labour market segmentation, and to identify further possible areas of research into the structure and operation of the labour market.

The application of both discriminant analysis and hierarchical classification analysis depends upon the variables on the specific occupations available for the analyses. The variables chosen for this thesis are of two types. The first are variables chosen from the statistics derived from the information provided in the household unit records and the personal records of the New England Region 1970 Public Use Sample.[2)] These variables are quantitative variables based upon continuous summary statistics of individuals grouped according to occupation. The second are judgemental variables. These variables are qualitative and refer to the general description of the occupation used by the U.S. Department of Commerce in creating the four broad occupational areas: White Collar, Blue Collar, Farm, and Service. The qualitative variables reflect a general description of the activities and the duties required in the job. They are sufficient to indicate broad job descriptions.

The choice of variables was first of all, limited by the amount of data provided in the 1970 Census. However, putting aside this initial limitation, we sought to include in the analyses as much information as possible without duplication and redundancy. In doing so, we sought

---

2) See Tables 5.3.1, 5.3.2, and 5.3.3 of Chapter 5, Section 5.3.

to represent in our variables the important social, institutional, and economic aspects of the labour market. At the same time we sought to represent in them subtleties of closely related aspects without an excessive multiplication of variables on any feature of the labour market. Stepwise discriminant analysis provides a safeguard against excess amounts of information and spurious results by omitting those variables which do not achieve a predetermined level of significance, that is, are not significant in the classification. Since stepwise discriminant analysis eliminates any unnecessary proliferation of variables and information, a greater emphasis was placed on the inclusion of as many relevant variables as could be derived from the New England Region 1970 Public Use Sample. The choice of the variables for the analyses covers as broad a range of the labour market as could be derived from the available data. The variables on each occupation are listed in Table 7.2.1.[3)]

7.3 Description of the Variables

Table 7.2.1 in Section 7.2 above lists all the variables which were created from our data to delineate as many social, institutional, and economic factors influencing the structure and the operation of the labour market as possible. We regard the variables as reflecting the significant economic, non-market, and non-competitive aspects of the labour market, and thus, indicating the fundamental dimensions of labour market segmentation. Some of the variables are demographic; whilst others relate to particular aspects of the labour market. The sixty variables that are listed in Table 7.2.1 are organised into the

---

3) For an explanation of any terms see Table 5.4.1 of Chapter 5 Section 5.4.

Table 7.2.1

Sixty Socio-Economic Variables Observed from each of the 428 Occupations

Notation	Variable Description
X1	Number of Males as a percent of the Labour Force
X2	Number of Females as a percent of the Labour Force
X3	Number of Whites as a percent of the Labour Force
X4	Number of Blacks as a percent of the Labour Force
X5	Number of Other Non-Whites as a percent of the Labour Force
X6	Number of Spanish-Americans as a percent of the Labour Force
X7	Median Age of Labour Force
X8	Median Education of Labour Force
X9	Number of Females with 1 or more children as a percent of all Females in the Labour Force
X10	Number of Individuals Married with Spouse Present as percent of the Labour Force
X11	Number of Individuals Married with Spouse Absent as a percent of the Labour Force
X12	Number of Individuals Never Married as a percent of the Labour Force
X13	Urban Population as percent of the Labour Force
X14	Rural Population as percent of the Labour Force
X15	Metropolitan Population as percent of the Labour Force
X16	Central City Population as percent of Labour Force
X17	Number of Persons Employed in the Central City of SMSA as a percent of Employed Persons
X18	Number of Persons Employed in the Ring of SMSA as a percent of Employed Persons
X19	Number of Persons Employed outside the SMSA as a percent of Employed Persons
X20	Number of Persons using Private Transportation to Work as a percent of Employed Persons
X21	Number of Persons using Public Transportation to Work as a percent of Employed Persons
X22	Number of Persons using Alternative Means to Work as a percent of Employed Persons
X23	Percent of Labour Force Employed in Agriculture, Foresting and Fisheries

Table 7.2.1 con'd

Notation	Variable Description
X24	Percent of Labour Force Employed in Mining
X25	Percent of Labour Force Employed in Construction
X26	Percent of Labour Force Employed in Manufacturing
X27	Percent of Labour Force Employed in Transportation, Communications, and Other Public Utilities
X28	Percent of Labour Force Employed in Wholesale/Retail Trade
X29	Percent of Labour Force Employed in Finance, Insurance, and Real Estate
X30	Percent of Labour Force Employed in Business and Repair Services
X31	Percent of Labour Force Employed in Personal Services
X32	Percent of Labour Force Employed in Entertainment and Recreation Services
X33	Percent of Labour Force Employed in Professional and Related Services
X34	Percent of Labour Force Employed in Public Administration
X35	Percent of Labour Force not Reporting an Industry
X36	Number of Employees of Private Enterprise as a percent of the Labour Force
X37	Number of Employees of the Government (Federal, State, Local) as a percent of Labour Force
X38	Number of Self-Employed as a percent of the Labour Force
X39	Number of Unemployed as a percent of the Labour Force
X40	Number of Employed who worked less than 26 weeks in 1969 as a percent of Employed
X41	Number of Employed who worked between 27 and 47 weeks in 1969 as a percent of Employed
X42	Number of Employed who worked over 50 weeks in 1969 as a percent of Employed
X43	Number of Employed who worked less than 14 hours per week as a percent of Employed
X44	Number of Employed who workes between 15 and 34 hours per week as a percent of Employed
X45	Number of Employed who worked between 35 and 40 hours per week as a percent of Employed

Table 7.2.1 con'd

Notation	Variable Description
X46	Number of Employed who worked over 40 hours per week as a percent of Employed
X47	Number of Persons Designated as Chief Income Recipient as a percent of Labour Force
X48	Median Total Personal Income
X49	Number of Persons Receiving Wages and Salary Earnings
X50	Number of Persons Earnings Positive Business Income as percent of Labour Force
X51	Number of Persons with a Loss in Business as a percent of Labour Force
X52	Number of Persons Earnings Positive Farm Income as a percent of Labour Force
X53	Number of Persons with a Loss in Farm Income as a percent of Labour Force
X54	Number of Persons Receiving Social Security Income as a Percent of Labour Force
X55	Number of Persons Receiving Welfare or Public Assistance as a percent of Labour Force
X56	Number of Persons Designated as below the Poverty Level as a percent of Labour Force
X57	Occupation described as White Collar Work
X58	Occupation described as Blue Collar Work
X59	Occupation described as Farm Work
X60	Occupation described as Service Work

following seven types of indicators of the labour market structure and conditions:

a. Barriers to Mobility and Economic Achievement
b. Family and Social Structure
c. Geographic Location
d. Industrial Structure
e. Work Climate and Patterns
f. Economic Performance
g. Institutional Structure

These seven types of indicators and the variables comprising each of the seven are described below within the context of the general concept of labour market segmentation. The purpose of this section is to provide a framework of labour market segmentation in terms of the sixty factors available in the analysis. This discussion of the sixty factors is presented in very general terms. It does not venture too far into the realms of sociology or social thinking. The discussion is intended to acknowledge the influences of these factors. Each set of indicators is considered in turn and discussed in isolation from the other variables. The discussion serves to point out the influences of each factor as though each were completely independently determined. This descriptive framework provides the basis of reference for the empirical work to follow in Chapter 8.

a. Barriers to Mobility and Eoonomic Achievement

The indicators of the institutional barriers to mobility and economic performance are intended to provide a view of serval aspexts of institutional mobility. The variables in this group of indicators include the extent of sexual, racial, and linguistic barriers to mobility. These variables and the level of educational achievement and age constitute the demographic characteristics of the labour force composition in the labour market segments. These variables are the composition of the labour force according to sex, variables X1 and X2; race, variables X3, X4 and X5; and Spanish-American population, variable X6.

On the one hand, labour market segmentation recognises institutional barriers to mobility and suggests that Women, X2, Blacks, X4, Other Non-Whites, X5, and Spanish-Americans, X6, will generally be concentrated in lower status and lower paying occupations. These factors influencing the composition of the labour market segment indicate a higher concentration of Women, Blacks, Other Non-Whites, and Spanish-Americans in the subordinate primary segment and the secondary segment. The variables, X2, X4, X5, and X6 should reflect the participation of the less preferred worker in the labour market. These variables should be less strongly associated with the index of the independent primary segment.

On the other hand, labour market segmentation suggests that males and Whites are preferred by the labour market. The variables of the male composition. X1, and the White composition, X3, of the labour force should be positively associated with the independent primary and subordinate primary segments of the labour market. However, males and Whites would tend to be preferred by the secondary segment as well, but the jobs in this segment are less attractive to the males and Whites. These members of the labour force will be relatively more concentrated in the higher status, and better paying occupations.

These demographic variables that indicate the institutional barriers to mobility do not account for all the differences in a person's position in the labour market. A substantial reason for Women, Blacks, Other Non-Whites, and Spanish-Americans being concentrated in the subordinate primary segment and the secondary segment is due to the direct effects of age and education. Age, variable X7, and education, variable X8, represent two additional variables reflecting institutional barriers to mobility, as well as to economic performance. Irrespective of the other demographic determinants, according to labour market segmentation teenagers and people over 45 should be associated with lower paying occupations; therefore, a concentration of these persons in the subordinate primary segment and the secondary segment occurs.

High education is more highly regarded and rewarded under labour market segmentation. Low levels of educational achievement are treated in just the opposite manner. Thus, very high levels of education, professional training, should be associated with the independent primary segment; moderate levels of education with the subordinate primary; and low levels of education with the secondary segment. Since Women, Blacks, Other Non-Whites, and Spanish-Americans generally receive less formal education in terms of number of years of schooling, they would face barriers to entry in the independent primary, and therefore, should be concentrated in the subordinate primary and secondary segments. Thus, these variables on the demographic characteristics and the variables on age and education should reflect relatively a greater non-male and non-White participation in the two lower segments of the labour market than in the independent primary segment.

b. <u>Family and Social Structure</u>

The family and social structure indicators reflect the aspects of familial, social, and cultural barriers to mobility and participation in the labour market. In terms of mobility, these indicators represent more a measure of social and cultural mobility than occupational mobility. However, the labour market may prefer certain characteristics of the labour force that likewise may be more generally accepted socially and culturally. Thus, to a degree. economic mobility and occupational mobility may adhere to the norm of the social structure.

The variables comprising this group of indicators are the number of women having borne children, and the marital status of the persons in the labour force. The significance of the women having borne children, variable X9, is realised when considering the type of work and the amount of work a woman can undertake. Whether a woman has children or not is a major factor influencing the participation of women in the labour market. The amount of work that a woman may be willing to undertake may to a large extent depend upon whether she has children, and vary with the number and ages of the children. Since the information on the ages of the children was not available, this variable simply indicates the number of women having given birth to children. If a woman has children, it should be more likely that she would undertake part-time or more flexible types of jobs in terms of committments. Occupations with part-time demand and variable levels of intensity and variable working time arrangements are more concentrated in the less bureaucratic labour market segments, the independent primary and the secondary segments. But, as females are more likely to be concentrated in the lower paying jobs; females having borne children should be more likely to be concentrated in the secondary segment than in the independent primary.

The marital status of persons in the labour force provides an indication of the familial and social pressures to achieve a satisfactory level of economic performance. The analysis in this thesis concentrates upon the way that the labour market operates in terms of the entire labour force. Therefore, the variables on marital status are taken over the entire labour force, not separately by sex. There are three variables on the marital status. The first is those married with the spouse present, variable X10; the second is those married with the spouse absent, variable X11; and the third is those never married, variable X12. Traditionally, these variables indicate the influences of the stability of family life and the committment to economic responsibility of the family. The married persons with the spouse present should tend to require a more stable income and stable employment to meet these committments and should be concentrated in the primary segments of the labour market. Indeed, even the employers of the subordinate primary segment should favour married persons with the spouse present because they provide a reliable source of stable, adherent labour. The married persons with the spouse absent do not bear the familial and social pressures to maintain stable and secure employment. Being designated with the spouse absent may even imply less rigid and less responsible attitudes, a characteristic undesirable to the employers of the subordinate primary. These persons should be associated with more flexible and less bureaucratic employment, that is, the structure of the independent primary segment and the secondary segment. The persons never married should be treated similarly by the employers in each of the labour market segments, and no tendencies for these persons to concentrate in one or the other of the labour market segments should exist.

c. Geographic Location

The indicators of the geographic location represent the geographic influences on the structure of the labour market. These variables serve to illustrate the influences of presence of different population size areas, the geographic organisation of employment, and the barriers to employment imposed by the means of transportation to work. This group represents the physical, geographic organisation of the economic system.

The first of these variables, variables X13 and X14 recognise the general geographic and population division between urban areas and rural areas, respectively. Urban areas are areas with 2500 or more inhabitants residing there. Variable X15 is a further stratification that identifies the number of persons residing within a standard metropolitan statistical area, SMSA. The most narrowly defined geographic residential area is presented as the number of persons residing in a central city of a SMSA, variable X16. All four variables, X13 - X16 are used to reflect the varying degrees of urbanisation of the sample area under consideration.

The development of the concept of labour market segmentation depended, in part, on the increasing specialisation of labour and of occupations. The specialisation of labour and of occupations is accompanied by a greater concentration of population. So, in highly urbanised areas there should be more highly skilled, higher paid, and higher status occupations. A high degree of urbanisation reflected in the variables X13 - X16 should be associated with occupations in the independent primary and subordinate primary segments. However, since areas of high urbanisation provide large housing markets and public transportation, the less preferred workers would seek out the low cost housing in these areas and would thus provide a natural supply of labour to the employers in the secondary labour market.[4] An area with a low degree of urbanisation should not exhibit the large hierarchy of jobs of highly urbanised areas; and therefore an area with a low degree of urbanisation should have more occupations in the secondary labour market segment.

The second of these variables specify the geographic location of employment. The variables X17, X18, and X19 indicate employment in the central city, in the ring of a SMSA, and outside a SMSA, respectively. They, then, reflect the degree of urbanisation and centralisation of the economic structure. The employment structure in the central city should be associated with the more skilled and better paid jobs of the primary labour market segments, as well as the less skilled, less preferred jobs of the secondary segment, necessary to maintain and to service the primary segments. Since large cities should generally tend to have a large supply of less preferred workers in residence, a source of cheap labour is available to the secondary segment. A high degree of urbanisation of employment should be associated with each of the three labour market segments. Employment located outside a SMSA should be less specialised because of the nature of industrialisation and should exist in the secondary labour market segment. The opportunities in the ring of a SMSA should be associated with the primary segments because of the general lack of public transport to reach the outlying areas of SMSA. Employment opportunities in these areas would require private transport which is usually unavailable to the less advantaged workers, the labour demanded by the employers of the

---

4) The housing markets include a broad range of housing (high cost flats to low cost, though not necessarily adequate, housing). Areas of high urbanisation generally tend to be characterised by areas of low cost housing, or put differently, cheap and inadequate housing.

secondary segment.

The third group of these variables provides added information on the location of employment, as well as on an aspect of geographic barriers to employment. These variables provide information on the persons using private means of transportation to work, X20, the persons using public means of transportation, X21, and the persons using alternative means of transportation, such as walking, bicycling, car pools, and others, X22. Persons using private means of transportation may require access to quick transport during work, prefer the luxury of private means, or have no alternative means of transport. Persons that rely on public means and alternative means of transportation may prefer using public transport and alternative means, have easy access to these means of transport, or may be forced to rely on public and alternative means out of economic necessity. These variables on the means of transportation indicate the nearness and accessibility of the location of employment to the location of residence. These variables serve to illustrate the relative importance of the dependence on transportation in obtaining and maintaining employment. The reliability and accessibility to different means of transportation will determine the geographic area within which to work and live and influence the maintenance of regular attendance and obeyance of working hours. The varying degrees of availability of the different means of transportation to the labour force represent an economic barrier to labour mobility and an economic limitation to opportunities for employment. Since the labour force of the secondary segment are lower paid and less able to afford private means of transport, they should show a greater reliance and use of public means and alternative means of transportation to work than private means. Conversely, the labour force of the primary segments should show little overwhelming preference. They may possibly show a reliance upon private means to enable them to commute from outlying residential areas to urban industrial areas. Often though, reliable public transport accommodates large numbers of these commuters from outlying residential areas.

d. Industrial Structure

This group of variables provides information on the way in which labour will be employed, or put differently, the type of labour demanded in the job. These variables indicate the industrial characteristics of the institutions and their control in the labour market.

The characteristics of the industrial structure in which the labour force are employed indicate the influences of industry on the treatment of labour in the labour market segments. One group of variables on the industrial structure is represented by variables X23 - X35. The variables, X23 - X35, depict the number of persons in an occupation that are employed with one of the respective industries: agriculture; forestry and fisheries; mining; construction; manufacturing; transportation; communications and other public utilities; wholesale and retail trade; finance, insurance, and real estate; business and repair services; personal services; entertainment and recreation services; professional and related services; and public administration.

The occupations located within the manufacturing; wholesale and retail trade; business repair services; finance, insurance and real estate; and public administration industrial sectors should be associated with stable working conditions, a stable environment, better pay, and established and routinised work patterns. These occupations should be concentrated in the subordinate primary segment.

However, higher level occupations within some of these industrial sectors and within the professional and related services sector are the most economically rewarding, the most highly skilled in terms of job requisites, and the most flexible, and should be associated with the independent primary segment. The agricultural, forestry and fisheries; mining; and service oriented (personal, entertainment, and recreation) sectors often exhibit conditions of low pay, less stable working conditions, and frequent turnover. The occupations in these industrial sectors should be concentrated within the secondary labour market segment.

The other group of variables in this category explains the influences of the mixed economic system on the labour market structure. The variables represent the level of employment in the private sector, X36, in the government sector, X37, and in the self-employed sector, X38. An occupation concentrated in government sector exists in a more favourable and stable atmosphere of employment. Fewer chances and risks of collapse are involved in government employment. The structure of government employment has well established and definite lines of bureaucratic procedure, lines of promotion, and a concentration of occupations at the middle level. In the private sector and the self-employed sector, there is more allowance for and recognition of initiative, creativity, and ambition. However, the majority of the occupations in these sectors exhibit the characteristics of the two lower segments of the labour market. A high concentration of employment in the government sector should be associated with the subordinate primary segment. No definite trend should be apparent from the private and self-employed sectors, except possibly a higher concentration of the occupations of the independent primary segment in these sectors rather than in the government sector.

e. Work Climate and Patterns

The variables which make up this indicator of the labour market structure are intended to refer to the rate of utilisation of the labour force, the pattern of employment, and the intensity of effort in the occupations. The unemployment rate, variable X39, signifies the demand for particular types of labour. The unemployment rate also indicates the overall level of economic activity in the economy. According to labour market segmentation, the labour force in the secondary labour market segment experience frequent job changes, and the jobs are usually considered the less preferable. The jobs in the primary segment are the more preferable and most demanded by labour. Barring any structural changes in the industrial composition of the economy, the occupations with the higher unemployment rates should be concentrated in the secondary segment. However, it is worth noting again that the labour force employed in the occupations of the independent primary segment also change jobs frequently and there may be a tendency to associate a higher unemployment rate with this labour market segment than with the subordinate primary segment.

The other variables of this indicator are of two kinds. One kind, variables X40 - X42, refer to the number of weeks worked per year; and the second, variables X43 - X46, to the number of hours worked per week. Both types of variables serve to illustrate the pattern of employment and the intensity of effort. If these variables indicate a stable routinised work pattern, that is a regular number of weeks worked per year, (48 to 52 weeks) and a regular number of hours per week (35 to 40 hours), the occupation should be associated with the subordinate primary labour market segment. If these variables show a less routinised activity pattern, the occupations should be associated with either the independent primary or the secondary segment. The occupations in the independent primary and the secondary segments should show a greater variability of activity than those in the subordinate primary segment. Of course, in the independent primary the variation in work habits and patterns is a matter of choice by the person; while in the secondary segment, this variation of activity level represents an inadequacy in the structure of employment offered, as well as a matter of choice in employment.

f. Economic Performance

The position of a person in any of the three labour market segments depends in part on whether a person may or may not be the primary income earner. The level of income varies greatly among the occupations and the labour market segments, and therefore, whether or not a person is the primary income earner will influence the motivation and ambition in seeking employment in the higher paying occupations. Whether a person is designated as the chief income recipient, variable X47, determines the economic necessity of the person to seek stable and better paying employment or, if not, to seek more flexible, less stable, and lower paying employment. If a person is the chief income recipient, there is a greater ecnomic necessity to maintain stable and better paying employment, and so, an association of these persons and their occupations with the primary labour market segments. There may be a tendency for an association of non-chief income recipients to be concentrated in the occupations of the secondary segment.

Variable X48, the total personal income, is the total income a person receives from all services and activities. It is an indication of a person's income status. The higher a person's total income, the more likely will the person be employed in an occupation of the primary segments. The persons with the highest income should be associated with the occupations of the independent primary segment. Moderate and middle level incomes should be associated with the occupations of the subordinate primary, and the lowest levels of income with those occupations of the secondary segment.

The group of variables X49 - X56 are intended to portray the income earning structure of labour market segmentation. Variable X49 refers to the number of persons receiving wages and salary income. The earning process of the subordinate primary and the secondary segment should be highly associated with earnings in the form of wages and salary. Persons with positive business earnings, X50, and business losses X51, could be associated with each of the three labour market segments. But these two variables should be more evident in the primary segments, as opportunities for self-employment are more limited to the labour force of the secondary segment. Those with positive farm earnings, X52, and farm losses, X53, should be associated with the secondary segment because the farm occupations in the New England Region are generally considered the lower paying. Persons receiving social security income, X54, welfare income, X55, and those persons designated as below the poverty level, X56, should be concentrated in the lower paying occupations. It is because these persons are employed in lower paying occupations that they receive social security or welfare and are classified as below the poverty level. Therefore, the occupations employing persons receiving public assistance should be associated with the secondary segment.

g. Institutional Structure

This category of indicators is intended to delineate in a broad and general manner the occupations according to the activities and the duties of the job. The variables represent general descriptions of occupations into White Collar, X57, Blue Collar, X58, Farm Work, X50, and Service Work, X60. The description of an occupation according to these variables implies different work conditions, relationships and institutional structure. White Collar work offers stable work, formal relations, rules of promotion, which should be associated with both the independent primary and the subordinate primary. However, the detailed descriptions are unavailable and specific conditions cannot be identified. If they were, creativity and initiative plus flexible work patterns would be evident in the independent primary, whereas, routinised work with strict adherence to rules would be associated with the subordinate primary. Blue collar work should be associated with both the subordinate primary and the secondary segment. The different occupations in the Blue Collar category will offer different conditions and structure of employment. Those with established approenticeship programs and high levels of union organisation should be associated with the subordinate primary; while the unstructured, unorganised, first-come first-hired occupations associated with the secondary segment. The occupations described as Farm Work or Service Work should be associated with both the subordinate primary and the secondary segment, but the association depends upon the work conditions and institutional structure. Since most occupations associated with Farm Work or Service Work are lower paying, are not effectively unionised, experience seasonal work, and exhibit little security, these occupations should be more highly concentrated in the secondary segment.

7.4 Selection of Sample

For the analysis in this thesis, we intend to include the broadest range of occupations possible. The list of occupational classifications for this analysis recognises the 428 unasterisked occupations listed in Appendix A. The most detailed level was selected because it provides the most detailed information on the social, institutional, and economic aspects of all the occupations that constitute the labour market segments. In other words, by concentrating on the detailed occupational classifications we intend to provide specific and detailed information on the aspects of the labour market. In doing so, the methods of analysis can rely on the specific and subtle similarities and differences among the occupations to provide a basis for the specification of the labour market segments. Thus, we can establish the composition, characteristics, and associations of the labour market segments with greater precision and clarity.

Although the method of hierarchical classification can be applied with any or all of the 428 occupations as the population for analysis, the technique of stepwise discriminant analysis cannot be performed with one group of occupations. Stepwise discriminant analysis must be based on some initial classification of the occupations into the three labour market segments. Thus, to perform the stepwise discriminant analysis, we must determine an initial sample of occupations for each of the three labour market segments.

In order to obtain the initial samples of occupations for each labour market segment, we relied upon some of the statistics on the occupations and subjective judgemental considerations based on the broad and general description of labour market segmentation. The sample of independent primary segment consisted of those occupations which showed a median level of income above $10,000 per year (X48) and, in addition, which showed most of the eight following characteristics: a high percentage concentration of males (X1); a low percentage concentration of women (X2); a high percentage concentration of Whites (X3); a low percentage concentration of Blacks (X4); a low percentage concentration of Spanish-Americans (X6); a median educational level of 16 years and above (X8); a high percentage concentration of employment in the professional and related industrial sector (X33); and a high rate of classification to the white collar description (X57).[5)]

The sample of the secondary segment included all those occupations which had a median income level below $5,000 per year (X48). In addition, the sample was selected to show a high relative percentage concentration with respect to five out of the ten following characteristics: Women (X2); Blacks (X4); Spanish-American population (X6); over 45 years of age (X7); median education below 11 years (X8); employment in the agriculture, forestry, and fishing industrial sector (X23); employment in the personal services sector (X31); poverty level income (X56); described as farm work (X59); and described as service work (X60). Many of these occupations showed earnings received from social security (X54) and welfare (X55).

Except for those occupations left unclassified, due to not being unambiguously assigned according to the characteristic described above, or to having fewer than ten individuals in the occupation, or to exclusion determined by subjective judgement, all the other occupations were included in the subordinate primary segment. This sample consisted of occupations in which the median income level was between $5,000 and $10,000 per year (X48) and a median education level between 12 and 16 years (X8). Many of these occupations showed a more equitable

---

5) These characteristics are not necessarily independent. For example, a high concentration of males (X1) implies a low concentration of women (X2).

concentration of males (X1) and women (X2) and concentration in both the White Collar (X57) and Blue Collar (X58) descriptions.

The sample groupings of the occupations into the three labour market segments and those left unclassified are given in Tables 7.4.1 - 7.4.4. The Tables 7.4.1 - 7.4.4 list the occupations in the independent primary segment, subordinate primary segment, secondary segment, and unclassified group, respectively. The four groups contain 57, 123, 115 and 133 occupations, respectively. The means of the sixty variables for each of the groups are presented in Table 7.4.5 and the variances of the sixty variables for each group are presented in Table 7.4.6. Thus, the discriminant analysis will be applied to the 295 occupations preliminarily classified into the three labour market segments.

By specifying these preliminary sample groupings of the occupations into the three labour market segments we are merely providing an initial classification on which to base a stepwise discriminant analysis of labour market segmentation. The samples provide sufficiently precise and inclusive groupings of occupations to fit the segment classifications according to the general description of labour market segmentation. It must be emphasises that this sample classification is only a starting point from which to apply the stepwise discriminant analysis. The sample classification need not be correct and must be considered with caution as subjective judgement that was beyond the realm of the statistics included in the analysis was used. However, the technique of stepwise discriminant analysis can accommodate misclassifications and can indicate the need for and provide further refinement of the classifications into the labour market segments.

Table 7.4.1

Preliminary Classification of

Occupations into the Independent Primary Segment*

Occupation Codes	
1	111
2	112
4	113
5	114
6	116
10	120
11	122
12	123
13	135
14	140
15	155
22	163
23	164
31	183
34	211
45	212
53	221
55	223
56	226
62	233
63	235
65	240
71	245
72	260
91	265
95	271
104	281
105	965
110	

* Occupations identified by occupation codes. To match an occupation code with its respective occupation, refer to Appendix A.

Table 7.4.2

Preliminary Classification of

Occupations into the Subordinate Primary Segment*

Occupation Codes			
3	213	426	522
25	215	430	523
32	216	431	525
36	220	433	530
42	222	435	533
82	224	436	535
100	225	440	543
124	230	441	545
141	231	445	552
142	270	452	554
143	282	455	561
144	284	456	575
150	285	461	601
151	301	462	631
152	305	470	640
153	312	471	641
154	313	472	652
161	321	473	661
162	323	475	680
171	326	480	703
173	331	481	704
181	334	482	705
184	343	484	712
191	361	485	714
192	363	492	715
194	390	495	802
201	412	502	953
202	413	506	961
203	415	510	962
205	420	514	964
210	422	516	

* Occupations identified by occupation codes. To match an occupation code with its respective occupation, refer to Appendix A.

Table 7.4.3

Preliminary Classification of

Occupations into the Secondary Segment*

Occupation Codes			
80	382	656	823
81	385	660	901
83	391	663	902
101	392	664	911
262	394	665	912
264	395	670	913
266	401	671	914
283	416	672	915
303	425	673	916
310	443	674	921
314	453	681	922
320	505	690	923
325	534	692	925
330	542	694	932
332	580	711	933
333	602	740	934
341	605	750	940
342	610	751	942
344	611	753	943
345	612	754	944
350	613	755	950
355	623	761	952
360	624	762	954
362	625	764	960
364	630	770	980
371	633	780	981
372	634	785	982
374	643	821	984
376	650	822	

* Occupations identified by occupation codes. To match an occupation code with its respective occupation, refer to Appendix A.

Table 7.4.4

Occupations Unclassified

Occupation Codes				
20	130	384	540	666
21	131	396	546	695
24	132	402	550	696
30	133	403	551	701
33	134	404	560	706
35	145	405	562	713
43	165	410	563	726
44	170	411	571	752
51	172	421	572	760
52	174	423	586	763
61	175	424	603	796
64	180	434	604	801
73	182	442	614	806
74	185	444	615	846
75	190	446	620	903
76	193	450	621	910
84	195	454	622	924
85	196	474	626	926
86	246	483	635	931
90	261	486	636	935
93	296	503	642	941
94	311	512	644	963
96	315	515	645	976
102	370	520	651	983
103	375	531	653	986
115	381	536	662	991
121				
125				
126				

* Occupations identified by occupation codes. To match an occupation code with its respective occupation, refer to Appendix A.

Table 7.4.5

Means of Sixty Socio-Economic Variables

Observed in the Preliminary Classifications

Variable	Independent	Subordinate	Secondary	Unclassified
1	84.96490	80.22742	41.58794	72.73047
2	15.03478	19.77176	58.41089	27.26889
3	97.63672	97.91618	95.83392	95.18526
4	1.27303	1.71508	3.79019	4.37341
5	1.08984	0.36794	0.37494	0.44062
6	0.84423	0.90671	0.98754	0.64743
7	41.98825	41.33263	38.78946	41.80580
8	17.44897	14.07370	12.87011	13.98783
9	41.84198	54.68387	60.93587	41.07042
10	78.90631	72.94614	52.58678	64.33176
11	6.37291	8.56895	14.57972	11.76053
12	14.72054	18.48369	32.83260	23.90686
13	68.47530	68.50784	68.89185	68.31949
14	20.73157	20.51990	18.98712	18.67918
15	47.12782	44.19197	43.75522	39.45853
16	19.76729	24.47556	28.18068	28.23964
17	30.85576	33.66396	36.19762	31.89209
18	27.87799	27.54234	28.21863	23.06387
19	41.26604	38.79306	34.71381	40.53209
20	83.69722	85.32576	70.16476	75.30336
21	6.83807	6.51342	9.98985	8.26311
22	9.46419	8.15959	18.97472	11.92150
23	2.64712	1.85647	3.50305	3.42486
24	0.06723	0.93737	0.02973	1.66638
25	1.65045	8.44866	3.67138	7.67772
26	23.68018	25.01649	30.07823	28.38527
27	5.71678	13.30679	3.34093	7.43903
28	3.53738	14.06433	17.62920	9.54413
29	7.58048	6.99564	3.55705	1.96981
30	2.63635	5.09649	3.07969	1.64234
31	1.79716	0.96908	10.26319	4.08402
32	0.13096	1.45194	2.07008	1.74840
33	44.01526	11.24363	18.37549	27.50735
34	6.54057	10.61281	4.40152	4.15852
35	0.0	0.0	0.0	0.75188
36	67.93556	72.58644	82.82010	74.77014
37	20.04385	19.49843	11.93726	15.97287
38	11.99158	7.74696	3.99549	8.85149
39	1.17244	1.79625	3.38707	3.39238
40	7.11254	11.21545	31.17110	17.49611
41	13.35570	13.28780	20.47783	17.91118
42	79.53131	75.49567	48.35019	62.33629
43	4.53366	3.65337	14.90922	6.75319
44	9.13367	8.75449	22.27715	15.11321
45	86.33209	87.59097	61.94310	73.62155
46	0.0	0.0	0.0	0.0
47	85.27493	74.79674	41.45291	68.24265
48	131.65179	74.11343	30.38893	64.60156
49	85.44017	85.85466	77.66521	81.95723
50	14.18843	7.02800	3.15597	7.29393
51	0.62074	0.36374	0.13593	0.70522
52	0.26884	0.45730	0.24106	1.12941
53	0.02125	0.10554	0.04825	0.35577
54	4.69696	8.46870	11.81651	12.19429
55	0.42584	1.13316	3.10590	1.36787
56	2.99243	3.51155	9.32839	6.62702
57	0.98246	0.46341	0.30435	0.42857
58	0.0	0.49593	0.43478	0.45865
59	0.0	0.00813	0.02609	0.02256
60	0.01754	0.03252	0.23478	0.08271

Table 7.4.6

Standard Deviations of the Sixty Socio-Economic Variables Observed in the Preliminary Classifications

Variable	Independent	Subordinate	Secondary	Unclassified
1	13.98907	23.31973	32.80560	32.81114
2	13.98908	23.31947	32.80559	32.81119
3	3.68809	2.71169	4.45043	13.56831
4	2.95232	2.58305	4.36452	13.49689
5	2.49160	1.07186	0.73153	1.91654
6	1.90300	1.77073	1.47445	1.86108
7	5.71229	6.07538	8.05030	9.80528
8	1.78528	1.58548	1.38758	2.89692
9	23.06938	32.86557	25.16438	37.94867
10	10.68005	13.92289	17.61188	26.78748
11	6.06559	6.48779	9.45414	18.39658
12	11.26304	11.69312	20.03104	22.56429
13	11.41537	11.51176	13.35282	22.81828
14	11.66841	9.99890	11.26049	19.24684
15	14.80901	10.61553	11.37326	23.17879
16	10.04025	8.91506	10.46637	18.89203
17	16.02158	13.23031	15.78277	22.61444
18	15.23407	10.97297	11.16602	19.08694
19	16.81839	12.89399	15.30407	27.21986
20	12.32412	10.84446	18.93652	28.34109
21	7.00274	8.54279	8.35194	16.69611
22	11.41305	8.46580	16.97752	19.14494
23	13.88204	10.50261	16.59471	17.23671
24	0.40758	9.02158	0.13411	10.19151
25	6.26375	18.91557	16.32779	22.09238
26	31.75301	30.77037	37.96945	37.52124
27	18.63586	28.40970	10.43847	21.99683
28	7.79334	24.84804	27.27490	21.08168
29	22.54361	20.80595	8.03688	9.82746
30	5.69334	10.51378	7.29361	5.44578
31	12.57687	4.30785	24.98801	17.20709
32	0.69659	8.88616	8.47446	8.96028
33	44.95680	23.79097	30.52553	40.07298
34	20.10493	25.91548	13.65152	15.82140
35	0.0	0.0	0.0	8.67109
36	28.33423	27.80150	22.46707	30.03473
37	25.09880	28.50728	20.07355	27.18756
38	23.13983	12.33881	9.79094	18.92477
39	2.02818	2.66033	3.02451	10.44604
40	6.15833	9.65633	16.21988	22.15424
41	12.15288	11.56090	8.36422	20.24280
42	14.85403	17.21011	16.78329	28.01494
43	8.45767	6.43992	17.91017	16.58463
44	10.67032	9.15095	12.03484	19.75766
45	15.50320	13.88676	23.13434	29.00604
46	0.0	0.0	0.0	0.0
47	9.55344	15.03390	16.66731	25.21095
48	30.36763	16.93932	10.42101	36.29645
49	19.00455	10.04629	12.18490	22.74315
50	19.47635	10.28738	6.50217	13.39908
51	1.33643	1.01195	0.49460	4.69707
52	1.18133	1.82694	0.93599	7.89889
53	0.10580	0.77626	0.27504	2.47227
54	4.81679	7.59729	9.37666	16.40399
55	1.04188	1.68866	3.18344	2.04786
56	4.01658	3.29221	7.29416	9.15253
57	0.13245	0.50070	0.46214	0.49674
58	0.0	0.50203	0.49789	0.50017
59	0.0	0.09017	0.16009	0.14905
60	0.13245	0.17810	0.42572	0.27648

7.5 Summary

The methods of analyses for clarifying the concept of labour market segmentation to be used in this thesis are stepwise discriminant analysis and hierarchical classification analysis. In using these two methods the analysis of labour market segmentation follows two approaches. The first is a conceptual approach that relies on a preliminary classification of a sample of occupations into the three labour market segments and the use of discriminant analysis. The second is an empirical approach that uses the technique of hierarchical classification to establish the initial classification of occupations into three groups. Discriminant analysis is then applied to this initial classification. Both approaches make possible estimation of classifications into three labour market segments. These approaches using the techniques of discriminant and hierarchical classification analyses produce statistical and empirical descriptions of the framework of labour market segmentation and call attention to the significant aspects contributing to the classification into one of the three labour market segments.

Both methods rely upon specific measured variables obtained from data on 428 occupations. Sixty variables are discussed in this chapter to represent the economic, non-market, and non-competitive factors of the social, institutional, and economic environment of the labour market. They summarise the important aspects of the structure, the conditions, and the operation of the labour market. The selection of the variables was limited by the type and amount of data available, and for this reason they do not cover all the aspects of the labour market which might appear to be significant in the contect of labour market segmentation.

In conclusion, before the stepwise discriminant analysis is performed, it is necessary to provide it with a first estimate of the three labour market segments. For this reason we have used a number of the objective characeristics defined in the sixty variables and subjective considerations of the occupations to establish a preliminary classification of 295 of the occupations into the three labour market segments. With this first approximation as a starting point we can now turn in Chapter 8 to the classification analysis of the concept of labour market segmentation.

CHAPTER 8

Classification Analysis of Labour Market Segmentation

8.1 The Study

The results of the classification analysis of labour market segmentation are presented in two sections. In the first section are the results of the conceptual approach in which we perform a discriminant analysis of the preliminary classification of a sample of the occupations into groups compatible with the concept of labour market segmentation. The second section contains the results of the empirical approach in which we firstly perform a hierarchical classification in order to group the occupations into three segments and, then secondly, a discriminant analysis of these segments.

To recapitulate, in both sections we apply the technique of stepwise discriminant analysis. This form of discriminant analysis selects from an initial list of variables those variables that will add most to the explanation of the differences between labour market segments given the other variable already selected. Variables are added one at a time to the discriminant analysis until no significant improvement in discriminating between the labour market segments can be obtained. There is one discriminant function associated with each labour market segment. The three discriminant functions establish the greatest possible separation between the segments and further ensure that the occupations are assigned to the most appropriate labour market segments.

8.2 Conceptual Approach

In the conceptual approach we used the technique of discriminant analysis to analyse the preliminary classifications of the sample of occupations that was presented in Chapter 7. We regard this preliminary classification as the best possible first approximation to the concept of labour market segmentation. Three sets of discriminant analyses were performed to analyse this approximation of the labour market segments. Each set of discriminant analyses used a different number of variables that served as the initial list of variables on which the analysis was based. The first set of discriminant analyses had an initial list of sixty socio-economic variables; the second set had an initial list of fifty-nine variables; and the third had sixteen variables.

a. Sixty Variables

The first discriminant analysis was performed with the initial list of sixty variables that was presented in Section 7.2 in Chapter 7. It is this list of sixty variables that provides the most comprehensive information with which to approximate the classification of all the occupations into the labour market segments and to describe the labour market segments. The stepwise discriminant analysis adds successively those variables that contribute to the explanation of the labour market segments. Variables with an F - statistic below .01 are excluded from the analysis.

Tests of Significance of the Discriminant Analysis

The usefulness of discussing the three labour market segments is enhanced by the several statistical tests of significance that are part of the discriminant analysis. Before turning to the discussion of the differences and similarities between the labour market segments and the performance of the discriminant functions, the ability of each variable

Table 8.2.1

Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE*	NUMBER OF VARIABLES INCLUDED
1	48		587.0740	1
2	47		45.3562	2
3	56		33.8208	3
4	8		30.1137	4
5	55		6.2006	5
6	26		4.2411	6
7	42		4.9965	7
8	20		3.5735	8
9	50		4.0103	9
10	1		3.0980	10
11	3		2.8976	11
12	40		2.3801	12
13	31		2.9221	13
14	32		2.3189	14
15	12		1.7832	15
16	5		1.3527	16
17	39		1.3848	17
18	29		1.3292	18
19	58		1.2989	19
20	36		1.3916	20
21	7		1.6223	21
22	27		1.1975	22
23	22		1.1031	23
24	6		1.0114	24
25	53		1.0093	25
26	23		1.0902	26
27	14		1.0571	27
28	34		1.1196	28
29	59		1.0073	29
30	16		1.0625	30
31	45		0.8744	31
32	52		0.8751	32
33	15		0.9363	33
34	49		0.8921	34
35	37		0.7749	35
36	38		1.2288	36
37	13		0.9526	37
38	33		0.7818	38
39	11		0.7064	39
40	24		0.7139	40
41	30		0.4357	41
42	54		0.3798	42
43	57		0.3929	43
44	9		0.3307	44
45	28		0.2512	45
46	17		0.1992	46
47	44		0.1819	47
48	51		0.0505	48
49	43		0.0340	49
50	19		0.0198	50

* F-ratio at .01 level with 50 and 243 degrees of freedom ≥ 1.60

selected for the discriminant analysis to discriminate between labour market segments, the significance of the differences between the means of each labour market segment, and the overall significance of the classifications were assessed.

These tests are summarised in Tables 8.2.1 and 8.2.2.[1] Table 8.2.1 lists the sixty variables in the order in which they were selected from the initial list of sixty for the discriminant analysis and their respective F - statistics. Significance of the F - value at the .01 level is accepted as evidence that a variable significantly discriminates between segments. Only sixteen of the sixty variables are significant at the .01 level. Even though not all the F - values on all sixty variables selected in the stepwise discriminant analysis was significant, it is necessary to point out that in order to obtain as much information as possible on the similarities and differences between labour market segments, we included in the analysis all variables with an F - value above .01.

Of the significant discriminations of the labour market structure, it is interesting but not unexpected that the median total personal income (X48) has the largest F - value and is the best discriminator of the sixty variables. The median age of the labour force (X7) has the lowest F - value and is the least effective discriminator of the significant variables. This suggests that the level of personal income is the single most reliable indicator of the labour market structure, whilst age of the labour force is the least reliable of the significant variables in discriminating between the labour market segments.

---

1) The descriptions of the variables are presented in Chapter 7.

Table 8.2.2

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	----------	12.477	30.394
Subordinate	12.477	----------	14.361
Secondary	30.394	14.361	----------
F - Statistic**	14.550		

* F-ratio at .01 level with 50 and 243 degrees of freedom ≥ 1.60

** F-ratio at .01 level with 100 and 486 degrees of freedom ≥ 1.41

Table 8.2.2 presents a matrix of F - statistics that can be used to test firstly whether each labour market segment is different from the two other segments, and secondly, whether on the whole the segments of the labour market structure are significantly distinct. In other words, the tests involve assessing whether the means of the labour market segments are significantly different from each other, and, in so doing, determining whether the classification depicts distinct groups in the labour market. The F - statistics in the F - matrix in Table 8.2.2 indicate that each labour market segment is significantly different from the other two. In addition, the F - statistic of 14.550 with 100 and 486 degrees of freedom indicates that on the whole the classification scheme represents three distinct labour market segments. Therefore, we can conclude in a general manner that the significance of the F - statistics in the F - matrix is due to the socio-economic variables listed in Table 8.2.1, and that this supports our contention of a labour market structure of three segments.

Discriminant Functions

The discriminant analysis used the sixty variables listed in Table 8.2.1 to generate one linear discriminant function for each labour market segment. The coefficients for the sixty variables in the three discriminant functions are presented in Table 8.2.3. [2] From

---

2) In Table 8.2.3 and the other tables similar to it, INDPRI means the independent primary segment, PRISEC means the subordinate primary segment, and SECSEC means the secondary segment.

Table 8.2.3
Discriminant Functions

VARIABLE	FUNCTION INDPRI	PRISEC	SECSEC
1	1.90861	1.86683	1.79531
3	16.74994	16.88028	16.67950
5	15.49100	15.19949	14.93550
6	-0.51079	-0.43489	-0.69149
7	5.79572	5.51203	5.45183
8	36.29625	34.84705	33.49467
9	0.09688	0.10676	0.10462
11	2.14173	2.20546	2.21304
12	0.96231	0.98094	1.02050
13	2.37266	2.35627	2.28852
14	2.58759	2.66114	2.61890
15	0.00150	0.02225	0.07615
16	2.59422	2.68651	2.69380
17	-0.59876	-0.62690	-0.62897
19	0.45570	0.44794	0.44335
20	-0.89823	-0.92385	-0.95301
22	-1.01395	-1.09119	-1.10610
23	-0.67144	-0.79027	-0.76881
24	2.01577	2.08916	2.09089
26	0.21211	0.19821	0.23048
27	0.67983	0.70001	0.68780
28	0.02089	0.04093	0.05141
29	0.57717	0.59463	0.57659
30	0.42616	0.39423	0.37922
31	1.26640	1.24597	1.22063
32	0.23510	0.22314	0.13901
33	0.23852	0.22678	0.23610
34	1.28550	1.25672	1.27953
36	5.23122	5.27089	5.46183
37	5.06576	5.18367	5.35916
38	5.72066	5.75265	5.91144
39	3.90201	3.75773	3.98091
40	-0.23152	-0.18933	-0.09759
42	1.07861	1.14014	1.16177
43	7.52000	7.54406	7.56333
44	7.40249	7.44100	7.48123
45	7.08947	7.15327	7.14175
47	-1.61288	-1.49791	-1.59055
48	-0.68313	-0.91828	-0.91993
49	-0.34839	-0.41688	-0.52390
50	-1.13731	-1.04411	-1.07652
51	-1.38243	-1.26824	-1.29355
52	-5.88050	-6.32789	-6.17721
53	-11.77134	-10.60310	-10.94529
54	-2.91292	-2.83337	-2.85161
55	7.87416	7.95080	8.13408
56	2.53022	2.35311	2.58423
57	-14.57011	-13.01349	-12.69601
58	38.82928	41.43546	42.42989
59	251.36453	265.86572	270.78003
CONSTANT	-1988.63306	-1961.20264	-1923.01782

these coefficients we can infer the following: 3)

(1) Six out of the eight variables that indicate the barriers to mobility and economic achievement were included in the discriminant analysis. The coefficient of the proportion of males in the labour force (X1) is the largest for the independent primary segment and smallest for the socondary. This means that given the other characteristics men are more likely than women to be employed in the primary labour market segments, and in particular in the independent primary. The coefficients for both for the proportion of Whites (X3) and other Non-Whites (X5) are higher for the primary segments than the secondary segment. The observation that Whites are more likely to be employed in the primary segments rather than the secondary is not unexpected, but the observaztion that the other Non-Whites (X5) are similarly treated by the labour market is not as expected. Furthermore, the coefficient for the number of Spanish-Americans (X6) is largest for the subordinate primary and smallest for the secondary. We would have expected the coefficient of this variable to be smallest for the indpendent primary segment.

The coefficients for age(X7) indicate that as the age of a person increases, the person is more likely to be classified in the primary segment than the secondary and most likely to be in the independent primary. This is as expected because the a priori expectation for the secondary segment was that it would be composed large numbers of teenage workers, as well as older workers, whilst the independent primary would more generally be composed of the older, more experienced, and more qualified workers. The coefficients for

3) The discussion to follow compares the results of the discriminant analysis with our expectations of the influence of each of the socio-economic variables as set forth in Chapter 7 Section 7.3.

educational attainment (X8) confirm our expectations that occupations requiring higher levels of education and persons with more education are more likely to be employed in the primary segments, and the highest levels of education would be required by the occupations in the independent primary segment.

(2) Three out of four family and social structure indicators were included in the analysis. The coefficient for women having borne children (X9) is nearly the same for the subordinate primary and secondary segments, thus a woman having borne children is less preferred by the independent primary segment than by the other two segments. The coefficients for both the percentage of individuals married with spouse absent (X11) and of the individuals never married (X12) are the largest for the secondary segment and the smallest for the independent primary segment. Even though our a priori expectation suggested that the coefficient for the independent primary should have been higher than that for the subordinate primary, the relative size and ranking of the coefficient for the secondary segment is as expected. The coefficient for both these variables indicates that occupations employing persons in these two categories are either subject to fewer and less demanding family and social responsibilities or to a form of social discrimination operating in the labour market or to a host of other possible reasons and are, therefore, most likely to be in the secondary segment in the labour market.

(3) The third type of indicator represents the influence of geographic location. Eight of ten of these were selected for the discriminant analysis. Variables X13 - X16 refer to place of residence. The coefficients for urban residence (X13) (that is a place of 2,500 or more inhabitants) indicate that occupations employing persons from an urban residence are more likely to be in the primary labour market

segments. However, this specification of urban residence includes city, as well as the more populated suburban, residences, and therefore does not incicate the influences of the different types of urban residences. Thus, we turn to the coefficients for both metropolitan residence (X15) (residence within SMSA) and central city residence (X16). The coefficients for both these variables are highest for the secondary segment and lowest for the independent primary. The relative rankings are as expected. The individuals employed in occupations in the secondary segment are most likely to live in the central and highly urbanized areas that tend to offer highly differentiated forms of employment and great variability in the supply of different types of housing. As expected, the coefficients for rural residence (X14) indicate that the persons in these areas are most likely to be employed in the subordinate primary segment possibly due to the economy being less structured and differentiated in these areas.

The coefficients for both employment in the central city (X17) and employment outside the ring of a SMSA (X19) are the highest for the independent primary and nearly the same for the other two segments. The coefficients for these two variables suggest somewhat contradictory results about the structure of employment in the independent primary segment. On the one hand, the highly professsional and specialised nature of the occupations in the independent primary require them to be located in highly urbanised and centralised areas that can supply the necessary support services. On the other hand, the jobs of this segment are generally characterised by more flexibility, creativity, and independence than the jobs in the other two segments. The occupations in the independent primary also offer more opportunities to carry on the activities of work in the less urbanised areas.

As expected the coefficient for the use of private transportation to work (X20) is highest for the independent primary and lowest for the secondary segment. In addition, the coefficients for the use of alternative means to work (X22) have the same relative rankings as for variable X20. For variable X20 this means that occupations in the independent primary either require more flexibility in the use of transportation for work or simply accommodate more suitably the financial needs for operating private transportation to work more than the other two segments. The results for variable X22 may reflect the fact that included in this factor are those persons that work at home and it is occupations in the independent primary segment that are most likely to allow this type of flexibility in working arrangements.

(4) The coefficients for the fourteen of the sixteen variables on the industrial structure used in the discriminant analysis generaly indicate features of the labour market structure not entirely expected. For one there is less difference between the relative sizes of the coefficients for these variables in each of the labour market segments than expected. This means that the different industrial sectors are characterised by occupational structures that cut across the labour market segments rather than each industrial sector being characterised by occupations predominantly grouped in a particular segment. For another we find that industrial sectors that are most likely to have occupations in the independent primary segment are also more likely to have occupations in the secondary segment than in the subordinate primary. The occupations of the mining (X24) and the wholesale-retail trade (X28) sectors are most likely to be in the secondary segment,

but unexpectedly, those of the manufacturing sector (X26) are also most likely to be in the secondary segment. While it is expected that occupations of both the professional services (X33) and the public administration (X34) sectors are most likely in the independent primary, it was not expected that more occupations of the agriculture, forestry, and fisheries (X23), and the business and repair services (X30), the personal services (X31), and the entertainment and recreation services (X32) industrial sectors would be in the independent primary than in the two other segments. The coefficients indicate that the occupations of both the transportation, communication, and other public utilities (X27) and the finance, insurance, and real estate (X29) sectors are most likely to be in the subordinate primary segments.

Furthermore, the coefficients for the three types of employment, that is private enterprise (X36), government service (X37), and self-employment (X38), are highest for the secondary segment and lowest for the independent primary. While we would have expected the weight for the subordinate primary to be the largest of the three segments for government employment and the weight for the independent primary to be larger than that for the subordinate for self-employment, they are not. Nevertheless, the coefficients for these variables serve to indicate that the three types of employment in the economy are characterised by occupations from all three labour market segments. The coefficients further indicate that there are more secondary segment jobs offered by the three types of employment than independent primary which is to be expected.

(5) The discriminant analysis used six out of eight of the variables on work climate and patterns. The first of these is

unemployment(X39). From the coefficients for unemployment we infer that unemployment is most likely to occur in the secondary segment. But, within the primary segments, unemployment is more likely in the independent rather than in the subordinate segment. This corresponds to our a priori expectations that less stable employment opportunities are more likely in the independent primary and secondary segments than in the subordinate primary.

The coefficient for percentage of the labour force that worked less than 26 weeks (X40) is highest for the secondary segment and lowest for the independent primary. We would have expected the coefficient for the independent primary to be higher than the subordinate primary. Even though we expected the coefficient for percentage of labour force that worked more than 50 weeks (X42) to be highest for the subordinate primary, it is not. It is highest for the secondary segment.

The number of hours worked per week is an important factor of the labour market structure. The coefficients for both percentage of those that worked less than 14 hours per week (X43) and of those that worked between 15 and 34 hours per week (X44) are highest for the secondary segment. Thus, the occupations in this segment offer employment on a basis less regular than a normal work week of 35 to 40 hours. As expected, the coefficient for percentage of persons that worked between 35 and 40 hours per week (X45) confirms our expectation that the occupations in the subordinate primary are associated with the most stable work week.

(6) All ten of the variables on economic performance were included in the discriminant analysis. The weights for the chief income recipient (X47) indicate that if a person is the chief income recipient,

the person is most likely to be in an occupation grouped in the secondary segment. Furthermore, the weights for the total personal income (X48) mean that as the level of income paid in an occupation increases, the probability that the occupation will be classified in the independent primary segment increases the most, whilst the likelihood that the occupation will be classified in the secondary segment increases the least.

The type of income offered influences the total level of personal income and, in turn, the labour market segment in which an occupation is classified. The coefficients for wages and salaries (X49) is highest for the independent primary and lowest for the secondary segment. For the variables on positive business income (X50), negative business income (X51), negative farm income (X53), and social security income (X54), the coefficients are highest for the subordinate primary and lowest for the independent primary. These results suggest that the occupations in the subordinate primary and secondary segments do not necessarily offer more alternative types of income, but in offering different types, the occupations do so with less reliable outcome in terms of increasing total income than those in the independent primary segment. One particular result of the coefficients for social security indicate that occupations in the subordinate primary offer the most stable work patterns that permit more of an opportunity for persons in these occupations to meet social security requirements in the event of unemployment to receive this form of income. As expected, the coefficients for both welfare income (X55) and being designated below the poverty level (X56) are highest for the secondary segment which implies that occupations in this segment are more likely to offer lower levels of compensation that require outside assistance to maintain subsistence.

(7) The last type of indicators represents the institutional structure. Three out of these four variables were included in the analysis. We would have expected the coefficients for White collar work (X57) to be highest for the independent primary and lowest for the secondary segment. The coefficients are just the opposite.[4)] However, the coefficientsfor this variable are large and negative and contribute the least of all the variables in discriminating between the labour market segments. The coefficients for both Blue collar work (X58) and Farm work (X59) confirm our a priori expectation that occupations described as either of these two are more likely to be in the secondary and subordinate segments than in the independent primary segments.

Performance of Discriminant Functions

The discriminant functions in Table 8.2.3 were used to obtain scores for each observation. The scores in the discriminant analysis used in this thesis are the Mahalanobis' D^2. This score represents the distance of an observation from the discriminant function used to obtain the score. An occupation is assigned to the labour market segment to which it is closest, that is, the smallest Mahalanobis' D^2.

The matrix in Table 8.2.4 illustrates firstly how many misclassifications occurred with the discriminant functions, and secondly how the unclassified occupations were assigned. From Table 8.2.4 we can see that a very small number of occupations in each of the preliminary classifications were considered misclassifications. However, the number of misclassifications obtained does not provide a completely satisfactory evaluation of the significance of the classifications.

The general picture observed in Table 8.2.4 can be complemented

---

4) We reexamine this variable and others in later analyses.

Table 8.2.4

Classification Matrix

Segment Before the Analysis	Predicted Segment: Independent	Subordinate	Secondary	Total
Independent	55	2	0	57
Subordinate	4	114	5	123
Secondary	0	3	112	115
Previously Unclassified	20	65	48	133
Total Classified	79	184	165	428

Table 8.2.5

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	57	18	31.6	11	19.3
Subordinate	123	65	52.8	12	9.8
Secondary	115	63	54.8	22	19.1
Total	295	146	49.5	45	15.3

* Chi^2 at .10 level with 50 degrees of freedom ≤ 37.50

** Chi^2 at .99 level with 50 degrees of freedom ≥ 76.10

by a test of the significance of the classifications. The test statistic is the Mahalanobis' D^2 that served as the discriminant score. The Mahalanobis' D^2 is distributed as chi-squared. On the one hand, significance of the D^2 below the .10 level is accepted as evidence that an occupation is on the left tail of the distribution. We use the ten percent level for the D^2 to identify these occupations as <u>core members</u>. The ten percent level was selected because it is not a serious problem to be considered a core member as opposed to a member. On the other hand, significance of the D^2 above the .99 level is accepted as evidence that an occupation is on the right tail of the distribution. We consider these occupations as <u>non-members</u> of the segment. Identifying an occupation as a non-member is more important than identifying members since it implies that the occupation is very probably not a member of the segment. Thus, selecting the .99 level reduces the likelihood of misclassification.

The significance of the classifications is summarised in Table 8.2.5. In total, 146 (49.5%) of those occupations in the preliminary classifications are core members of the labour market segment to which they were preliminarily assigned. Separately, the occupations associated with the subordinate primary and those with the secondary segment appear to be the most cohesive with, respectively, 52.8% and 54.8% of the occupations identified as core members. In contrast, 45 (15.3%) of the occupations are non-members of the labour market segments. The subordinate primary has the fewest non-members at 9.8%, and the independent primary the most at 19.3%. These values seem sufficient to conclude that a good first approximation has been made.

Another means of examining the relationships among the labour market segments is by using the canonical variables and their coefficients to project a set of points in two dimensions. Each point corresponds to one occupation. Thus, we have a two-dimensional representation of the labour market segments. Figure 8.2.1 shows the grouping patterns of the occupations in a two-dimensional space.[5] An examination of Figure 8.2.1 reveals that there is considerable overlap between the composition of the segments, but that, generally there is dichotomy among the groups. Most of the occupations in the independent primary (I) are on the lower left of Figure 8.2.1, whilst of those in the secondary segment (S) are on the lower right. The occupations of the subordinate primary (P) dominate the upper middle region of Figure 8.2.1. From this we can conclude that the scores for the occupations in the independent primary and the secondary segments cover a similar range for the canonical variable on the vertical axis. This similarity can be associated with the similar coefficients in the discriminant functions for the variables on the industrial structure and the work climate and patterns. This corresponds to the a priori expectation that the independent primary and the secondary segments have a number of structural characteristics in common.

Refinement of the Labour Market Segments

Considering the overlap between the segments observed in Figure 8.2.1, the unexpected relative rankings of some of the coefficients for some variables, and the preliminary nature of our initial classifications, we decided to repeat the discriminant analysis. In order to obtain more compact labour market segments, we carried out three more discriminant analyses. In the second discriminant analysis we omitted those occupations that were identified as possible misclassifictions in our first analysis.

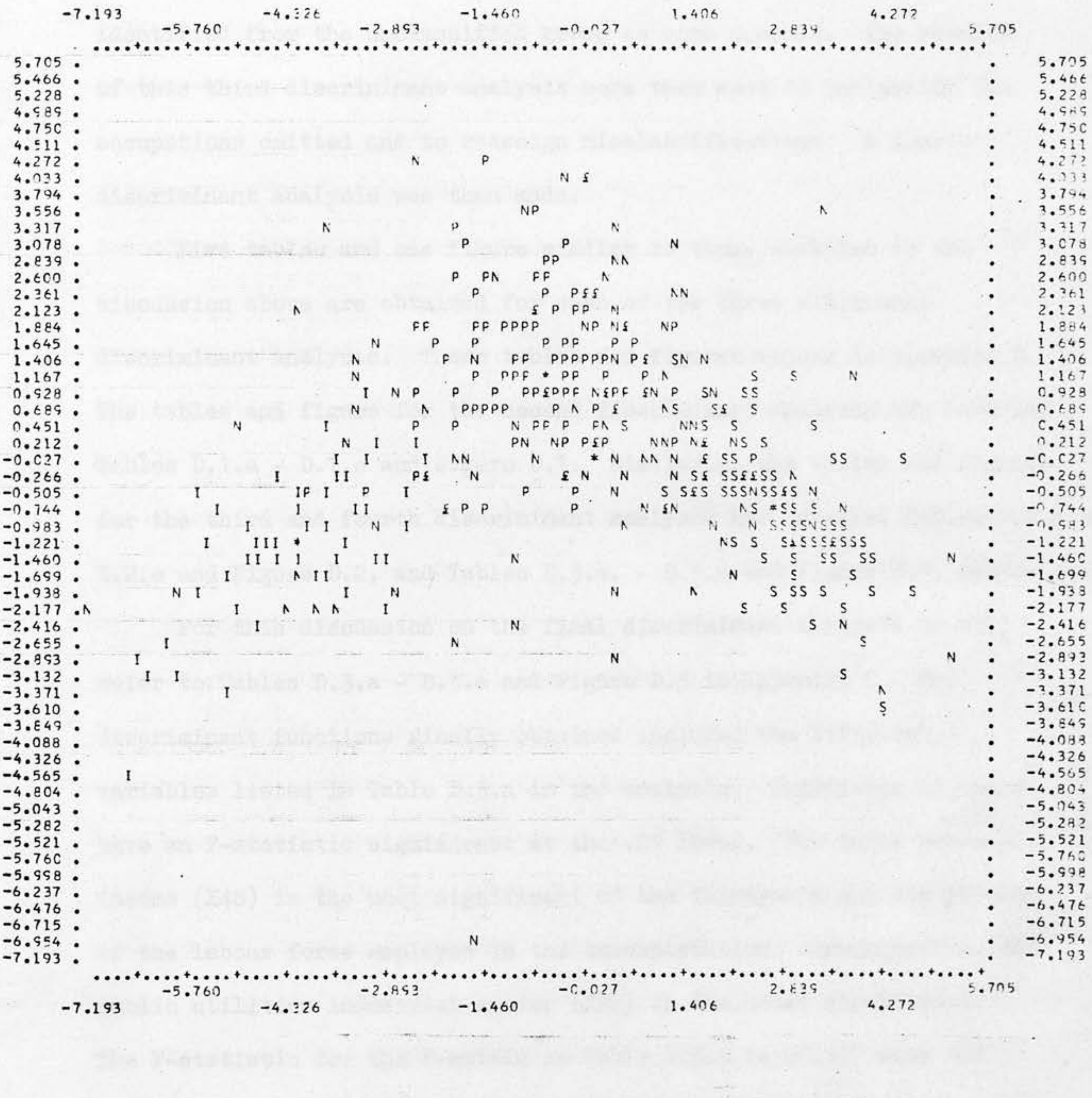
---

5) In Figure 8.2.1 and others similar to it, I = the independent primary segment, P = the subordinate primary segment, S = the secondary segment, N = not classified in the initial groups, * = group means, and £ or $\not{s}$ = point of overlap.

Figure 8.2.1

Two-Dimensional Representation

of the Labour Market Segments



The results of this second analysis were then used to identify misclassifications and non-members in the segments. In addition, we identified those occupations in the group left unclassified that would be significant core members of one of the segments. We repeated the discriminant analysis omitting the misclassifications and non-members and adding to the respective labour market segments the occupations identified from the unclassified group as core members. The results of this third discriminant analysis were then used to reclassify the occupations omitted and to reassign misclassifications. A fourth discriminant analysis was then made.

Five tables and one figure similar to those analysed in the discussion above are obtained for each of the three additional discriminant analyses. These tables and figures appear in Appendix D. The tables and figure for the second discriminant analysis are labelled Tables D.1.a - D.1.e and Figure D.1. Similarly, the tables and figures for the third and fourth discriminant analyses are labelled Tables D.2.a - D.2.e and Figure D.2, and Tables D.3.a. - D.3.e and Figure D.3, respectively.

For this discussion on the final discriminant analysis we will refer to Tables D.3.a - D.3.e and Figure D.3 in Appendix D. The discriminant functions finally obtained included the fifty-two variables listed in Table D.3.a in the analysis. Thirty-six of these have an F-statistic significant at the .01 level. The total personal income (X48) is the most significant of the thirty-six and the percent of the labour force employed in the transportation, communication, and public utilities industrial sector (X27) is the least significant. The F-statistic for the F-matrix in Table D.3.b is 20.577 with 104 and 748 degrees of freedom and serves to indicate that the final labour market segments are distinct.

Of the fifty-two variables in this final discriminant analysis, forty-four are the same as those used in the first discriminant analysis. Six new variables were substituted for six variables in the first analysis. The proportion of Blacks (X4) was substituted for the proportion of Whites (X3); the percentage of individuals married with spouse present (X10) were substituted for the percentage of those married with spouse absent (X11); employment in the ring of a SMSA (X18) replaced employment in the central city (X17); not reporting an industry (X35) replaced the public administration sector (X34); percentage of labour force who worked between 27 and 47 weeks per year (X41) replaced percentage who worked less than 14 hours per week (X43); and description as Service work (X60) replaced description as White Collar (X57). The two other variables added are the percentage of persons using public transportation to work (X21) and employment in the construction industry (X25).

The coefficients of the proportion of Blacks (X4) confirms our expectation that Blacks are most likely to be employed in occupations that are in the secondary segment and least likely in those in the independent primary. In fact, the proportion of Blacks is the fourth most significant of the discriminators between the labour market segments. In contrast to the first discriminant analysis, in this analysis Spanish-Americans (X6) are, as in our a priori expectations, more likely to be in those occupations in the secondary segment rather than in the subordinate primary. The coefficients for percentage of labour force married with spouse present (X10) indicate that if a person is married and the spouse is present, the person is more likely to be employed in an occupation in either the subordinate primary or secondary segment rather than the independent primary. This adds to our belief that the subordinate primary encourages more stable life styles.

The coefficients for employment in the ring of a SMSA (X18) indicate that if an occupation is located in the ring of a SMSA, it is most likely to be in the secondary segment and least likely to be in the independent primary, but we would have expected the reverse. From the weights of the use of public transportation to work (X21), we can conclude that those in the independent primary are most likely to use public transport. This could be due to the number of persons employed in the independent primary who live outside a SMSA and commute by public transport into the SMSA for work.

The coefficients for the employment in the construction industry (X25) confirm our expectations that it is most characterised by occupations in the subordinate primary and least characterised by those in the independent primary. In the first discriminant analysis we found relatively substantial differences in the coefficients for the variables on the type of employment (X36 - X38), but in this discriminant analysis we find that the coefficients are much more close in size. From these coefficients we can conclude that each of the types of employment is characterised nearly equally by occupations in all three segments.

The inclusion of percentage of labour force who worked between 27 and 47 weeks per year (X41) strengthened the association between the independent primary and secondary segments. The coefficients on this variable indicate that those occupations in these two segments are more likely to exhibit less stable work patterns than those in the secondary segment. This observation coupled with our previous conclusions on variables X39 and X40, confirms our a priori expectation that the occupations in the independent primary and secondary segments are characterised by less routinised and less secure work environments than those in the subordinate primary.

The coefficients for Service work (X60) indicate that this description is most characteristic of the work in the secondary segment. Furthermore, this description of occupations as Service type work is more characteristic of those in the independent primary than those in the subordinate primary because of a larger relative concentration of occupations in the independent primary in the professional services and personal services industrial sectors.

In the discriminant functions there are two unexpected results. The first is that the coefficient for occupations described as Farm work (X59) is higher for the independent primary than for the secondary. However, this may be due to the increasing centralisation and organisation of large scale farming that requires a greater influx of highly skilled managers relative to the addition of farm labour. The other is that the coefficients for percentages of persons designated as below the poverty level (X56) is largest for the independent primary segment and smallest for the secondary. We would have expected the opposite. The only explanation is that a number of occupations and the persons employed in them, although with a poverty level of income, have considerably more characteristics similar to those in the independent primary rather than to those in the other two segments.

For this final discriminant analysis we used an initial classification of 79 occupations in the independent primary, 186 in the subordinate primary, and 163 in the secondary. From the test of significance of the classifications in Table D.3.e in Appendix D, we find that 254 (59.3%) of the occupations are core members and that 67 (15.7%) are non-members. The most compact and distinct of the segments is the subordinate primary in which 119 (63.9%) are core members and only 22 (11.8%) are non-members. The 15.7% that were non-members of the classifications are due to the degree of heterogeneity of the segments. Nevertheless, the size of the proportion of core members (59.3%) seems high enough to indicate that this final classification is a reasonable representation of the occupations in the labour market segments. The final classification after this discriminant analysis assigns 78 to the independent primary, 186 to the subordinate primary, and 164 to the secondary. The occupation codes of the occupations in the three labour market segments are listed in Tables 8.2.6 - 8.2.8 respectively. Finally, the two-dimensional representation of the segments shown in Figure D.3 in Appendix D clearly illustrates the distinct nature of the groups. Again as in the first discriminant analysis, the points associated with the occupations in the independent primary and secondary segments span a similar range of values for the canonical variable on the vertical-axis. This illustrates the similarity between these two segments on certain variables that make-up this canonical variable.

b. Fifty-Nine Variables

In each of the four discriminant analyses performed with the initial list of sixty variables, total personal income (X48) had the largest F-statistic of all the variables. In other words, total personal income was the variable that most successfully and significantly discriminated between the labour market segments. However, the fact that high levels of pay are associated with the primary segments and the highest levels of pay with the independent primary was not unexpected. In fact, the preliminary classification of the sample of the occupations depended largely upon the level of personal income. It would seem only logical, therefore, that the discriminant analysis would reflect the

Table 8.2.6

Occupations in the Independent Primary Segment**

Occupation Codes**		
2*	91*	164
4*	94	165
5*	96*	174
6*	102*	183*
10*	103*	202*
11*	104*	211*
12*	105*	212*
13*	110*	221
14*	111*	223*
15*	112*	226*
22*	113*	233*
23*	114*	235*
30*	115*	240*
31*	116*	245*
34*	120*	260*
43	121*	265*
45*	122*	271*
52	123*	281*
53*	124*	411
55	125*	434
56	126*	546
61*	132*	550
62*	135*	713
63*	140*	801
65*	155*	802
71*	163*	965

** Occupations identified by occupation codes. To match an occupation code with its respective occupation, refer to Appendix A.

* Occupations identified by all three sets of discriminant analysis in the conceptual approach.

Table 8.2.7

Occupations in the Subordinate Primary Segment**

Occupation Codes*					
1	170	284*	436*	510*	635*
8	171*	285*	440*	512	640*
20	172*	312*	441*	514*	641*
21	173*	313*	442*	515*	642*
25	181	315*	444*	516*	651*
32*	182	321*	445*	520	652*
33	184	323*	446*	522*	653*
35	185	326	450*	523*	662*
36	190	331*	452*	525*	666*
42	191*	334*	454*	530*	680*
44*	192	343	455*	531*	701*
51	193*	361*	456*	533*	703*
64	194	363	461*	535*	704*
82*	195	374*	462*	540*	705*
86	196*	381	470*	543*	706*
90	201*	390*	471*	545*	712*
93	203*	402*	472*	552*	714*
95	205*	403*	473*	554*	715*
100*	210*	404*	475*	560*	726*
130	213*	412*	480*	561*	761
133	215*	413*	481*	563*	770*
141*	216	415*	482*	571*	903*
144*	220*	420*	483*	575*	910*
150	222*	422*	484*	586*	924
151*	224*	423*	485*	601*	935*
152*	225	424*	486*	603*	940
153*	230*	426*	492*	614*	961*
154*	231*	430*	495*	615*	962*
161*	246*	431*	502*	620*	963*
162*	270*	433*	505*	622*	964*
24	282*	435	506*	631*	562*

** Occupations identified by occupation codes. To match an occupation code with its respective occupation, refer to Appendix A.
* Occupations identified by all three sets of discriminant analysis in the conceptual approach.

Table 8.2.8

Occupations in the Secondary Segment**

Occupation Codes*					
72	311*	395*	624*	696*	915*
73	314*	396*	625*	711	916*
74*	320*	401*	626*	740*	921*
75*	325*	405*	630*	750*	922*
76	330*	410	633*	751	923*
80*	332*	416*	634*	752*	925*
81*	333*	421*	636*	753*	926*
83*	341*	425*	643*	754*	931*
84*	342*	443*	644*	755*	932*
85*	344*	453*	645*	760	933*
101*	345*	474*	650*	762*	934*
131	350*	503*	656*	763	941*
134	355*	534*	660*	764*	942*
142*	360*	536	661	780*	943*
143*	362*	542*	663*	785*	944*
145	364*	551*	664*	796*	950*
175*	370*	572	665*	806	952*
180*	371*	580*	670*	821*	953*
261	372*	602*	671*	822*	954*
262*	375*	604*	672*	823	960
264*	376*	605*	673*	846*	976*
266*	382*	610*	674*	901*	980*
283*	384*	611*	681*	902*	981*
296*	385*	612*	690*	911*	982*
301*	391*	613*	692*	912*	983*
303*	392*	621*	694*	913*	984*
305*	394*	623*	695*	914*	986*
310*					991*

** Occupations identified by occupation codes. To match an occupation code with ist respective occupation, refer to Appendix A.

* Occupations identified by all three sets of discriminant analysis in the conceptual approach.

importance of this factor as a discriminator of the labour market structure.

Furthermore, in our discussion of the a priori expectations of the influences of the sixty variables we noted that a number of the variables affect the earning power of individuals and that the variable on total personal income may reflect the influence of these variables. In fact, from the correlation matrix in Table D.4 in Appendix D we find that total personal income (X48) is positively correlated with a number of the variables that significantly discriminate between the segments. These variables include the proportion of males (X1), the proportion of Whites (X3), the level of educational attainment (X8), the proportion of labour force married with spouse present (X10), and proportion of labour force designated as chief income recipient (X47).

Therefore, we decided to repeat the discriminant analysis omitting total personal income from the initial list of variables. By doing so, we hoped that more explanatory power would be attributed to the remaining fifty-nine variables. In addition, we wanted to determine whether the classifications were more than just an ordering of occupations according to income size. In other words, if the classifications of occupations achieved in the first set of discriminant analyses with an initial list of sixty variables is a reasonable representation of labour market segmentation, then performing a similar series of discriminant analyses on an initial ist of fifty-nine variables should generate a similar classification.

The procedure for this set of four discriminant analyses is the same as that described and used in the first set of discriminant analyses with an initial list of sixty variables. Just as for the first set of discriminant analyses, the first discriminant analysis with a list of fifty-nine variables is performed on the preliminary classification presented in Chapter 7. The second, third, and fourth discriminant analyses use the results of the preceding discriminant analysis to obtain further refinements of the classifications. The fourth discriminant analysis produces the final discriminant functions and classifications.

As in the first set of discriminant analyses, five tables and one figure were obtained for each of the four separate discriminant analyses performed in this set. These tables and figures appear in Appendix E. The tables and figure on the first discriminant analysis are labelled Tables E.1.a - E.1.e and Figure E.1. The set of tables and figure on the second, third, and fourth discriminant analyses are the sets in Appendix E labelled E.2, E.3 and E.4, respectively. In the following discussion in this subsection we refer to Tables E.4.a - E.4.e and Figure E.4 in Appendix E.

The discriminant functions obtained by this procedure included the fifty-one variables listed in Table E.4.a. Thirty-eight of these variables have an F-statistic significant at the .01 level which is evidence that these thirty-eight are significant discriminators of the labour market structure. It is interesting that chief income recipient (X47) and the level of educational attainment (X8) have the two largest F-statistics and are, thus, the most significant discriminators. The four next largest F-statistics are for the proportion of Whites (X3), employment by the government (X37), age (X7), and the proportion of males (X1). These four variables were among the ten most significant discriminators listed in the final discriminant analysis using an initial list of sixty variables. On the whole, since the F-statistic of 20.583 with 102 and 750 degrees of freedom for the F-matrix in Table E.4.b is significant at the .01 level, these labour market segments are also considered different one from another.

The final discriminant functions in this set are presented in Table E.4.c. Forty-six of the variables used in the final analysis in this set are the same as those used in the final discriminant analysis of the first set. Obviously, total personal income (X48) is not one of the variables included in this analysis. The five additional variables used are the following; the proportion of Whites (X3), employment in the central city (X17), employment in the public administration industrial sector (X34), working week less than 14 hours (X43), and occupation described as White Collar (X57).

The coefficients for education (X8) show greater relative size differences in these discriminant functions than those in the final discriminant analysis with an initial list of sixty variables. The relative ranking is the same, that is, the highest levels of education are associated with the independent primary segment and the lowest with the secondary segment. The coefficients of chief income recipient (X47) indicate that a person designated as chief income recipient is most likely to be employed in an occupation in the subordinate primary. The coefficients of most of the forty-four other variables that were the same in this analysis as in the final analysis of the first set allow us to draw conclusions similar to those in subsection 8.2.a above.

The inclusion of variables X3 and X57 are important in the assignment of an occupation to the independent primary segment. The coefficients for both variables are the largestfor the independent primary and smallest for the secondary. Therefore, high values for these variables indicate a greater likelihood of being classified in the independent primary than in the two other segments.

Employment in the central city (X17) adds significantly to the information on the influence of the geographic distribution of occupations. The coefficients of this variable indicate the occupations in the secondary segment are most likely to be located in the central city. The coefficients for the public administration sector (X34) are relatively small and close in size and do not indicate any strong influence. The last of the five additional variables is percentage of labour force who worked less than 14 hours per week (X43). The closeness in size of the coefficients for this variable in the independent primary and secondary segment coupled with a similar pattern of the coefficients for the other variables reflecting the work climate and pattern indicate that more variable work patterns are more common to both the independent primary and secondary segments than to the subordinate primary.

The performance of these discriminant functions is summarised in Tables E.4.d and E.4.e. The classification used to obtain these functions had 91 occupations in the independent primary, 170 in the subordinate, and 167 in the secondary segment. Of the 428 occupations, 258 (60.3%) are core members and 65 (15.2%) are non-members. These figures are nearly the same as those found in the test of the significances of the classifications for the final discriminant analysis using an initial list of sixty variables. Just as in the final analysis of the first set, the most cohesive segment is the subordinate primary in which 114 (67.1%) are core members and only 15 (8.8%) are non-members. This is an improvement over the significance of the composition of this segment in the final analysis in the first set. However, the classifications in the other two segments are less significant in this discriminant analysis than in the former final analysis. The final classification has 88 occupations in the independent primary, 171 in the subordinate primary, and 169 in the secondary. The two-dimensional representation in Figure E.4 illustrates that the labour market segments are relatively distinct.

One reason we performed this set of analyses with an initial list of fifty-nine variables was to determine whether the classifications were more than merely a ranking by income size. If in omitting total personal income (X48) from the analysis, the final classifications are similar to those obtained in the analysis using all sixty variables, then we can consider the classifications to represent a multi-dimensional ranking of occupations into segments reflecting the concept of labour market segmentation. Table 8.2.9 presents the final number of occupations assigned to each segment by the two sets of discriminant analysis and further presents the number of occupations in each respective segment obtained in the final analysis using fifty-nine variables that were also assigned to the respective segments in the final analysis using sixty variables. From Table 8.2.9 we can conclude that a very high proportion of the occupations in the respective segments obtained in the final analysis using fifty-nine variables are the same as those obtained in the final analysis of the first set. Therefore, the labour market segments can stand as a classification scheme that goes beyond a one-dimensional ranking according to income size, and that in the absence of income data, occupations and individuals can still be assigned with minimum possibility of misclassification to one of the labour market segments.

c. <u>Sixteen Variables</u>

We decided to repeat the procedure of disciminant analysis again in order to determine if a similar final classification of the

Table 8.2.9

Comparison of Final Classifications

Discriminant Analysis	Number of Occupations		
	Independent	Subordinate	Secondary
Sixty Variables	78	186	164
Fifty-Nine Variables	88	171	169
Occupations Common to Both	66	156	158

occupations into labour market segments could be obtained, but with less information on the occupations. Thus, we set up the discriminant analysis with an initial list of sixteen variables. These are the same sixteen used to establish the preliminary classification of the sample of occupations. The same procedure of discriminant analysis used in subsections 8.2.a and 8.2.b above was used in this series of discriminant analyses using an initial list of sixteen variables. In other words, four different discriminant analyses were performed in order to arrive at a final classification on the basis of information on the sixteen variables. Five tables and one figure were obtained for each of these four discriminant analyses. The tables and figures appear in Appendix F. There are four sets of tables and figures in Appendix F with the set labelled F.1 corresponding to the first discriminant analysis, F.2 to the second, F.3 to the third, and F.4 to the fourth analysis. In the following discussion of the final discriminant analysis in this set, we refer to Tables F.4.a - F.4.e and Figure F.4 in Appendix F.

Table F.4.a lists the fifteen variables selected for the final analysis. Of these fifteen, ten had F-statistics significant at the .01 level. The variable with the largest F-statistic was total personal income (X48). Although much less successful than variable X48 as discriminators of the labour market structure, the proportion of males (X1) and the level of educational attainment (X8) have F-statistics nearly the same and are equally important in discriminating between segments. For the 428 occupations and the list of sixteen variables, the F-statistic in the F-matrix in Table F.4.b of 56.917 with 15 and 441 degrees of freedom is significant at the .01 level.

The coefficients for the discriminant functions are presented in Table F.4.c. The coefficients for most of the variables allow us to draw similar conclusions as we did in the discussion in subsection

8.2.a above. However, the coefficients for the proportion of Whites (X3) is highest for the secondary segment and lowest for the independent primary. Although the relative rankings of the coefficient for this variable are different than expected, the relative sizes of the coefficients are quite similar. This nearness in size could be due to the large proportion of Whites relative to the other races in the sample. In the analysis using all sixty variables, the coefficients for being designated below the poverty level (X56) were not as expected. But, in this analysis the coefficients are as expected, that is, highest for the secondary segment and lowest for the independent primary. The other unexpected result in this analysis is the order of the relative ranking of the coefficients for the occupation description White Collar (X57). We would have expected the independent primary to be highest and the secondary the lowest, but in this analysis the positions are reversed. Nevertheless, this is not totally contradictory as a number of White Collar jobs are low paying and lower level dead-end jobs, such as receiptionists, mail boys, mailing clerks, and others.

The classification in the final analysis had 77 occupations in the independent primary, 187 in the subordinate primary, and 164 in the secondary. The evidence on the significance of these classifications in Table F.4.e is not as acceptable as that on the analysis using sixty variables. Only 147 (34.3%) of the occupations are core members and 44 (10.3%) are non-members. The compactness of the subordinate primary improved markedly over that in the analysis using sixty variables, but at the loss of cohesiveness of the other two segments. The final classification obtained in this analysis assigns 78 occupations to the independent primary, 194 to the subordinate primary, and 156 to the secondary segment. The two-dimensional representation of the segments shown in Figure F.4 indicates that the

Table 8.2.10

Comparison of Final Classifications

Discriminant Analysis	Number of Occupations		
	Independent	Subordinate	Secondary
Sixty Variables	78	186	164
Sixteen Variables	78	194	156
Occupations Common to Both	63	163	149

distinctness between the segments is less defined than in the two-dimensional representation for the final classifications in the analysis using sixty variables.

Table 8.2.10 presents the number of occupations in the final classifications obtained by both the discriminant analysis using sixty variables and that using sixteen variables and the number of occupations in common to the respective segments in both analyses. From Table 8.2.10 we can conclude that even with as little information as is provided on the sixteen variables used in this set of analyses, we can assign occupations to labour market segments, but as noted by the evidence from the test of the significance of the classifications, with less certainty and less probability of minimising the chance of mis-classification.

d. Comparison of the Three Sets of Discriminant Analyses

To compare the final classifications for each of the three sets of discriminant analysis, we present Table 8.2.11. This table summarises the number of occupations classified to the labour market segments in each of the three sets of analysis and presents the number of occupations in the respective segments commonly identified in all three sets of analysis as being a member of that segment. Of the 428 occupations, 353 (82.5%) were assigned to the same labour market segment by three sets of discriminant analysis that used different amounts of information on the occupations to perform the classifications. The occupation codes listed in Tables 8.2.6 - 8.2.8 that have asterisks are those occupations obtained for the respective segments by all three sets of discriminant analysis.

Table 8.2.11

Comparison of Final Classifications
Obtained in the Three Sets of Discriminant Analysis

Discriminant Analysis	Number of Occupations		
	Independent	Subordinate	Secondary
Sixty Variables	78	186	164
Fifty-Nine Variables	88	171	169
Sixteen Variables	78	194	156
Occuaptions Common to All Three	61	146	146

8.3 Empirical Approach

The empirical approach is an analysis conducted to determine whether the classifications used in the conceptual approach were just constructions to fit a concept or whether the concept actually accommodates the existence of segments in the labour market. In other words, if labour market segments do exist, the empirical approach should generate groups of occupations that are characterised by similar attributes to those described in the conceptual approach.

The method of analysis adopted in this empirical approach was to perform a hierarchical classification with the occupations to establish classes of occupations. With these classes, we then employed the technique of discriminant analysis to analyse the variables used for the classifications and to assign any misclassifications or unclassified occupations. The use of discriminant analysis enables us to assess the significance of the variables, the significance of the distinctness of the classes, and the quality of the classifications.

As a reminder, in the technique of hierarchical classification the occupations are classified into groups on the variables provided with the Euclidean distance as the measure of dissimilarity and Ward's algorithm as the grouping procedure. In a sense, each occupation is represented by a point in as many dimensions as there are values on the variables of the labour market structure. The greater the distance between the points in the space, the more dissimilar the occupations are. Those two occupations that are the closest in the space are combined to form a new group or point. The procedure continues until all the occupations are combined into one group. At any stage of the classification, the groups have the greatest possible homogeneity. However, with each combination, the degree of dissimilarity increases.

In the empirical approach, two different hierarchical classifications were performed. The first was performed on the total population of 428 occupations, whilst the second was performed on a sample of 295 occupations. With the results of each hierarchical classification, we applied the technique of discriminant analysis to supplement the analysis of the classifications obtained by the hierarchical classifications.

a. Hierarchical Classification on 428 Occupations

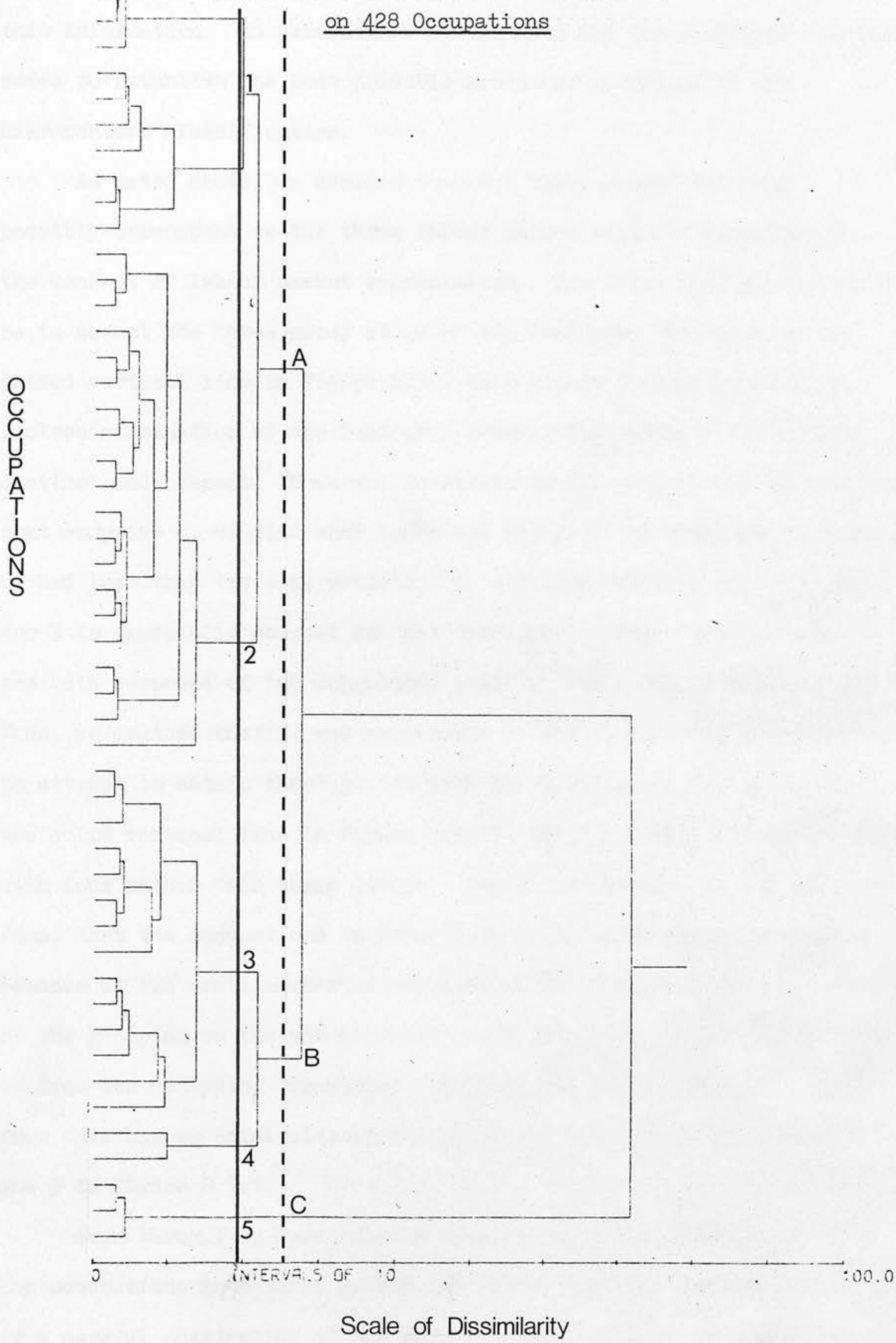
In the first hierarchical classification, each of the 428 occupations in the population is characterised by its values on the sixty variables of the labour market structure. It is the scores on these sixty variables that determine which occupations are nearest to each other in the labour market structure. Since we want to consider a labour market with three segments, it was decided that the most meaningful classification is one that identifies three groupings of occupations. Thus, a grouping would be designated as either the independent primary, the subordinate primary, or the secondary segment.

The results of this first hierarchical classification are illustrated in the dendogram in Figure 8.3.1. The dendogram in Figure 8.3.1 is not the complete dendogram from 428 to 1 grouping, but only that for the groupings of occupations from 50 to 1 group. The scale at the bottom of the dendogram shows the level of dissimilarity for each grouping. Thus, the dendogram illustrates the value of similarity or dissimilarity within each group and the increase in the dissimilarity created by each successive grouping.

From the hierarchical classification shown in Figure 8.3.1 we would like to obtain three relatively distinct and compact groups. In other words, we want three groups that have a high level of

Figure 8.3.1

Dendogram: Hierarchical Classification on 428 Occupations



dissimilarity between one another and the lowest level of dissimilarity among members of the groups. The dendogram summarises pictorially all this information. An examination of the data and the dendogram can then serve to establish the best possible groupings generated in the hierarchical classification.

As noted above, we decided to select three groups that may possibly correspond to the three labour market segments described in the concept of labour market segmentation. The first inclination would be to accept the three group stage of the dendogram indicated by the dotted vertical line in Figure 8.3.1 that yields the groups A, B, C. A close examination of the dendogram reveals that group C is the most distinct and compact. However, in examining the data on the occupations that comprise C, we find that these are fifty of the occupations on which we had less than ten observations. In addition, we find that neither A nor B is especially compact at this three group stage. Groups A and B are both composed of two subgroups, 1 and 2, and 3 and 4, respectively. Thus, we decided that it was preferable to omit C from the analysis and to attempt to obtain three groups from the group stage indicated by the solid vertical line in Figure 8.3.1. Nevertheless, this leaves us with four rather than three groups. Again, we referred to the data and found that the occupations in Group 4 is composed of seven farm occupations. Because of the small number of occupations in this group and the nature of the grouping on the characteristics of farm work, we decided to omit it from the analysis. Therefore, the best set of groupings we selected from this hierarchical classification is the classification scheme 1, 2 and 3 in Figure 8.3.1.

Even though, we were relying upon an empirical classification of the occupations into three groups, we cannot discount the importance of a careful examination of the data and the dendogram in attempting to optimise the groupings. It is interesting to note that hierarchical classification identified the distinct group C that happens to be composed of those occupations on which we had less than ten observations. Since the remainder of the analysis in this empirical approach is heavily dependent upon the initial classifications of occupations, it was necessary to carry out a thorough examination of the data and the dendogram in order to obtain the classifications 1, 2 and 3, in Figure 8.3.1.

We have tentatively called group 2 the independent primary, group 3 the subordinate primary, and group 1 the secondary segment. The usefulness of the three groups is suggested by the observation that the means and standard deviations for the groups summarised in Tables 8.3.1 and 8.3.2, respectively, are sufficiently different. Even though the relative sizes of the means for the income figures for these groups do not correspond exactly with those figures for our preliminary classifications developed in Chapter 7, the relative rankings are the same. In addition, the relative rankings for all the variables across the groups do not correspond exactly with our a priori expectations of labour market segments. But, as this is an empirical approach divorced from subjective influences, the rankings do not necessarily have to be those identified in our a priori expectations. The purpose of this approach is to establish to what extent the occupations classified in the segments by the empirical approach correspond to those in the respective segments obtained in the conceptual approach.

A comparison of this initial classification obtained in the hierarchical analysis on the 428 occupations with the preliminary classifications obtained in Chapter 7 is summarised in Table 8.3.3. In Table 8.3.3 we find that only the group that we call the secondary

Table 8.3.1

Means:

Hierarchical Classification on 428 Occupations

VARIABLE	GROUP INDPRI	PRISEC	SECSEC	NEWGRP
1	64.51018	78.71877	48.84354	77.24252
2	35.48894	21.28020	51.15590	22.75731
3	97.70625	96.59221	96.03754	94.29892
4	1.46322	3.10615	3.50092	5.33647
5	0.82984	0.30071	0.46041	0.36447
6	0.42989	1.35392	0.73123	0.40419
7	40.77055	41.22958	39.72940	42.02145
8	15.92099	12.83463	13.94502	14.73825
9	48.23567	55.29877	57.50894	29.58006
10	67.11298	69.14697	56.82585	66.73584
11	8.75813	10.10825	13.90213	12.50924
12	24.12788	20.74339	29.27151	20.75485
13	68.73186	67.74992	72.09459	64.63943
14	20.21802	19.40611	17.17720	22.56358
15	44.76395	41.78117	47.09886	36.09396
16	23.26476	27.36482	28.62210	23.58411
17	35.19173	33.21848	37.66309	23.42960
18	26.11183	27.33487	29.06601	19.82880
19	37.84137	39.44513	33.27063	46.21512
20	80.03862	84.40828	72.51814	64.06409
21	7.33505	5.83752	11.50142	10.01881
22	11.77067	9.75283	15.97950	15.39078
23	1.52243	0.99920	1.06017	14.03509
24	0.00898	1.04950	0.04467	2.98246
25	1.30555	12.91808	1.60266	3.42105
26	13.96589	43.24269	20.42892	20.43858
27	8.59967	9.36889	5.71652	5.08772
28	11.19725	13.44746	13.99446	7.92398
29	8.99122	1.10317	6.81734	1.75439
30	1.20246	4.75206	4.41538	0.58479
31	1.08255	2.63231	13.49051	2.45614
32	1.02720	1.37811	2.57627	1.31579
33	40.99826	6.66268	21.33318	31.75438
34	10.09816	2.44506	8.51942	6.49123
35	0.0	0.0	0.0	1.75439
36	66.10786	83.78090	77.59177	67.18939
37	24.65787	8.68265	16.43135	21.52617
38	9.10036	7.20395	5.67551	9.13551
39	1.78916	3.11668	1.94734	4.14468
40	14.75225	15.93209	25.15515	18.70288
41	16.65276	16.63625	16.10176	17.69370
42	68.59404	67.43048	58.74236	58.34019
43	8.13079	4.68038	13.08827	6.99060
44	12.61958	13.19014	18.56517	14.78398
45	78.39395	82.12825	68.34555	67.69905
46	0.0	0.0	0.0	0.0
47	66.56866	68.27650	54.94077	70.41093
48	82.13071	62.55371	53.20399	71.55005
49	81.86858	83.91856	81.93901	79.86211
50	9.60365	7.01435	4.43624	5.97929
51	0.57374	0.18294	0.23587	1.24616
52	0.54483	0.18125	0.12634	2.54088
53	0.02597	0.21013	0.01988	0.49719
54	7.39565	9.89697	10.61913	14.80412
55	0.72160	2.31580	2.25185	0.63263
56	4.28209	6.16292	8.02578	5.56162
57	0.90598	0.14907	0.51613	0.47368
58	0.08547	0.83230	0.09677	0.33333
59	0.0	0.0	0.0	0.12281
60	0.00855	0.01863	0.38710	0.05263

Table 8.3.2

Standard Deviations:

Hierarchical Classification on 428 Occupations

VARIABLE	GROUP INDPRI	PRISEC	SECSEC	NEWGRP
1	30.99001	25.81317	35.13907	35.24361
2	30.99002	25.81326	35.13921	35.24362
3	3.61617	3.90072	5.20480	19.77898
4	3.19275	3.86261	5.00178	19.69933
5	2.02606	0.87874	0.69831	2.64920
6	0.87943	2.36310	0.93281	1.69543
7	6.93861	6.33343	7.92523	12.76253
8	2.35268	1.71538	1.87241	3.04947
9	31.17476	36.55139	19.78677	40.03107
10	16.77696	17.21735	20.49396	34.98181
11	6.71574	7.26500	11.09856	26.36218
12	17.45930	15.41338	18.56079	28.15717
13	13.25327	10.70182	9.53659	33.62796
14	12.25785	9.13330	7.50788	28.80002
15	13.02168	11.21103	10.35637	32.74405
16	11.20549	9.10342	10.43372	26.27740
17	16.18800	12.03774	15.44147	29.86766
18	14.31823	11.07942	9.04386	26.02013
19	17.58409	13.40096	11.86895	37.69699
20	15.69507	10.58448	17.85576	39.96426
21	9.06294	5.99551	8.51690	24.30524
22	12.64120	8.56767	15.98724	29.37186
23	6.52839	8.11790	5.34761	35.04375
24	0.06867	8.31003	0.15035	14.87620
25	9.18065	25.93033	2.94948	15.15270
26	28.18723	38.52046	22.93748	37.27634
27	26.74046	22.09292	13.32244	19.74139
28	27.24213	22.02568	19.45557	23.52446
29	25.04630	6.30315	10.18699	13.24525
30	3.14137	10.88996	6.14331	4.41505
31	8.80100	11.21768	29.10266	13.66438
32	8.66240	7.22573	7.99168	9.93394
33	44.69565	17.40193	26.94804	45.37021
34	27.31079	6.49881	19.18367	23.33687
35	0.0	0.0	0.0	13.24532
36	31.33087	17.33899	22.98199	40.12175
37	31.36998	13.67175	22.16563	36.31635
38	18.99814	11.84854	12.66243	23.03687
39	3.32486	3.52181	1.77862	15.25815
40	12.79226	13.62117	18.84914	30.65167
41	13.67729	9.55696	8.63466	29.05803
42	20.59923	18.45070	22.86620	38.30875
43	13.96332	7.89153	18.26326	20.32596
44	11.33218	11.30124	13.43542	27.76483
45	21.33104	16.17242	24.85133	39.55145
46	0.0	0.0	0.0	0.0
47	23.98927	19.31961	22.47992	33.43086
48	43.73273	28.78094	37.17703	50.33977
49	16.04686	10.91256	12.10214	32.50015
50	16.22495	10.65449	8.27114	13.93207
51	1.47610	0.53788	0.49380	7.08688
52	2.00472	0.50907	0.41095	12.02269
53	0.17337	1.98610	0.09383	2.11844
54	7.79092	7.78136	9.26018	23.15746
55	1.27394	2.60161	2.79387	2.05266
56	5.54844	5.18790	8.11666	11.39531
57	0.29311	0.35726	0.50245	0.50374
58	0.28078	0.37476	0.29725	0.47559
59	0.0	0.0	0.0	0.33113
60	0.09245	0.13565	0.48973	0.22528

Table 8.3.3

Comparison of Initial Classifications:

Hierarchical Classification on 428 Occupations

Empirical Approach	Conceptual Approach		
	Independent	Subordinate	Secondary
Independent	35	37	21
Subordinate	6	64	44
Secondary	14	18	46

segment has a composition distinctly similar to our preliminarily classification of occupations as secondary segment. The group we called the subordinate primary is composed of a larger number of subordinate primary occupations than of the other two segments. However, the group called the independent primary contains nearly equal numbers of independent primary and subordinate primay occupations. We regard these results as sufficient evidence for continuing to identify the three hierarchical classification groups with the segment to which we tentatively assigned them.

In this approach, we then perform one discriminant analysis of the classifications because of the greater statistical certainty in which the classifications were achieved. The three groups to be used in the discriminant analysis contain 117, 161 and 93 occupations and 57 are left unclassified. The discriminant analysis selected forty-nine variables from the initial list of sixty. These variables are listed in the order in which they were selected in Table 8.3.4.Of these forty-nine, twenty-four have F-statistics significant at the .01 level and are, thus, successful discriminators of the labour market structure. The two most important discriminators are the descriptions as either White Collar (X57) or Blue Collar (X58). The most successful discriminator in the conceptual approach was total personal income (X48) which is not introduced in this analysis until the 29th step. The evidence summarised in the F-matrix in Table 8.3.5 on the significance of the groups being different one from another is significant at the .01 level and sufficient to conclude that the classifications are viable.

Forty-six of the forty-nine variables in the discriminant functions listed in Table 8.3.6 are the same as the variables used in the first set of analyses in the conceptual approach. The coefficients

Table 8.3.4

Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE*	NUMBER OF VARIABLES INCLUDED
1	58		232.1968	1
2	57		55.5195	2
3	52		13.9394	3
4	33		14.9974	4
5	6		15.8032	5
6	30		7.3659	6
7	2		6.2704	7
8	49		7.3006	8
9	55		6.9994	9
10	25		4.2599	10
11	26		5.2258	11
12	21		3.0461	12
13	50		3.0698	13
14	51		3.5935	14
15	5		2.3786	15
16	40		2.4109	16
17	20		2.6687	17
18	39		2.1272	18
19	27		2.1163	19
20	18		2.1784	20
21	47		2.3082	21
22	23		1.6851	22
23	12		1.8305	23
24	15		1.4951	24
25	14		0.9283	25
26	10		1.0024	26
27	22		0.8397	27
28	8		0.9037	28
29	48		1.8921	29
30	29		0.9020	30
31	4		0.8211	31
32	32		0.8992	32
33	53		0.7676	33
34	7		0.7316	34
35	9		0.6520	35
36	17		0.6680	36
37	44		0.3402	37
38	13		0.2962	38
39	54		0.2530	39
40	34		0.2437	40
41	37		1.3141	41
42	36		0.3415	42
43	38		0.7958	43
44	24		0.2085	44
45	16		0.1576	45
46	31		0.1315	46
47	56		0.1082	47
48	45		0.0357	48
49	41		0.0258	49

* F-ratio at .01 level with 49 and 320 degrees of freedom ≥ 1.63

Table 8.3.5

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	----------	14.361	6.094
Subordinate	14.361	----------	8.841
Secondary	6.094	8.841	---------
F - Statistic**	9.414		

* F-ratio at .01 level with 49 and 320 degrees of freedom ≥ 1.63

** F-ratio at .01 level with 98 and 640 degrees of freedom ≥ 1.43

Table 8.3.6

Discriminant Functions

VARIABLE	FUNCTION INDPRI	PRISEC	SECSEC
2	0.96809	0.98905	1.01480
4	2.75807	2.76637	2.69355
5	4.95782	4.50416	4.64238
6	8.88569	9.65698	9.35892
7	9.02553	8.95650	8.96544
8	30.94273	30.73180	30.36267
9	0.26964	0.27905	0.27988
10	5.23906	5.19285	5.20092
12	6.11070	6.03908	6.09746
13	2.26844	2.24189	2.27163
14	0.16488	0.10048	0.11459
15	-0.94738	-0.92296	-0.90065
16	0.06754	0.07165	0.05475
17	-1.84072	-1.86694	-1.86317
18	1.58352	1.56343	1.59564
20	2.95631	3.02813	2.94295
21	2.50365	2.61769	2.56217
22	3.65634	3.70823	3.63081
23	-4.80449	-4.82374	-4.75126
24	-3.65835	-3.68245	-3.65989
25	-1.98765	-1.93622	-1.92811
26	-2.48202	-2.44727	-2.44447
27	-1.54167	-1.52818	-1.52212
29	-1.64308	-1.66280	-1.64908
30	-3.20487	-3.08186	-3.03571
31	-1.58043	-1.57268	-1.57278
32	-2.34420	-2.32659	-2.36279
33	-1.46100	-1.50179	-1.48619
34	0.27218	0.22666	0.25876
36	426.81665	426.29980	426.22876
37	425.05664	424.57886	424.48511
38	430.06177	429.53638	429.50146
39	3.38691	3.29656	3.18206
40	1.24390	1.31354	1.30627
41	-0.61364	-0.61676	-0.61978
44	3.23689	3.26457	3.25141
45	2.37575	2.38079	2.38220
47	-1.19819	-1.17009	-1.15696
48	-0.97571	-0.96099	-0.94556
49	0.25248	0.37178	0.32788
50	-3.84357	-3.73559	-3.83812
51	-12.04774	-12.71221	-12.29936
52	4.34468	3.44074	3.91644
53	414.30713	414.15161	413.97607
54	-0.67963	-0.70937	-0.68143
55	0.30069	0.63156	0.56815
56	1.19294	1.19841	1.21694
57	-0.67185	-4.17889	-7.27903
58	51.01711	55.68918	44.71790
CONSTANT	-22181.50391	-22141.07812	-22119.58203

of twenty-four variables allow us to infer similar conclusions to those presented in section 8.2. However, for a number of variables, such as proportion of Blacks (X4) and total personal income (X48), the rankings of the coefficients are at variance with our a priori expectations of the structure of the labour market. But, the coefficients for other variables, such as the description as either White Collar (X57) or Blue Collar (X58), confirm our a priori expecations with greater certainty in terms of coefficient size than appeared in the conceptual approach.

For other variables that did not appear in the discussion in section 8.2, the coefficients correspond to our expectations. For example, the coefficients for the proportion of women (X2) is highest for the secondary segment and lowest for the independent primary. This means, as expected, that occupations in the secondary segment offer more likelihood of employment to women than the other two segments. Most of the results of this discriminant function complement our explanations of the labour market segments. Nevertheless, the explanation for some of the results, such as those on the proportion of Blacks (X4) and total personal income (X48), can only be explained by the heterogeneous composition of the groups as illustrated in Table 8.3.3.

The classification matrix shown in Table 8.3.7 indicates that more misclassifications were made in this analysis than in the analyses of the conceptual approach. Furthermore, in the test of the significance of the classifications presented in Table 8.3.8 we find that only 176 (47.4%) of the 371 occupations in the initial classification were core members and 61 (16.4%) were non-members. This evidence is less significant than that on the classifications in the conceptual approach. The most distinct and compact of these groups is the secondary segment and the least distinct is the independent primary. Figure 8.3.2, the

Table 8.3.7

Classification Matrix

Segment Before the Analysis	Predicted Segment Independent	Subordinate	Secondary	Total
Independent	101	3	13	117
Subordinate	11	136	14	161
Secondary	13	8	72	93
Previously Unclassified	26	18	13	57
Total Classified	151	165	112	428

Table 8.3.8

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	117	33	28.2	27	23.1
Subordinate	161	83	51.6	20	14.3
Secondary	93	60	64.5	14	15.1
Total	371	176	47.4	61	16.4

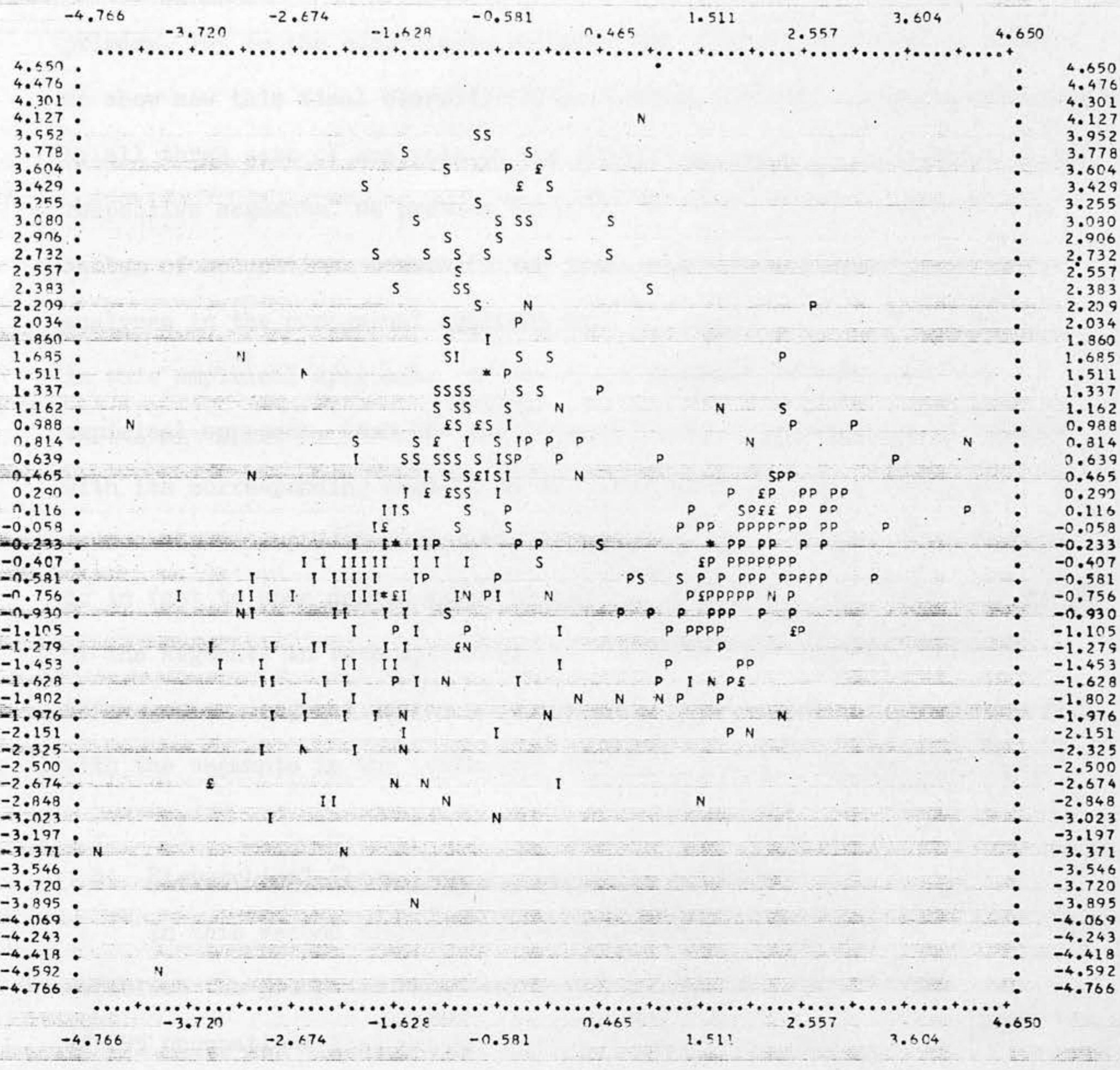
* Chi^2 at .10 level with 49 degrees of freedom $\leq$ 36.70

** Chi^2 at .99 level with 49 degrees of freedom $\geq$ 75.64

Figure 8.3.2

Two-Dimensional Representation

of the Labour Market Segments



two-dimensional representation of these groups, illustrates a considerable amount of overlap between the groups, that is, I's, P's, and S's are more intermingled within the periphery of the three groups.

The final classification assigns 151 occupations to the independent primary, 165 to the subordinate primary, and 112 to the secondary segment. To show how this final classification compared with the occupations identified by all three sets of analysis in the conceptual approach as members of the respective segments, we present Table 8.2.9. This table summarised the number of occupations common to the three segments obtained by the three analyses in the conceptual approach that are classified in the segments in this empirical approach. In the three segments obtained in this empirical approach, each has the largest number of occupations in common with its corresponding segment in the conceptual approach. This is reasonable evidence that segments depicted by labour market segmentation do in fact to some degree exist in the labour market. The most similar of the segments in both approaches is the secondary segment. The most heterogeneous segment in the empirical approach in terms of correspondence with the segments in the conceptual approach is the independent primary.

b. Hierarchical Classification on 295 Occupations

In this second hierarchical classification we decided to establish the groups in the labour market segments using the sample of 295 occupations that was selected for the preliminary classification in Chapter 7. In addition, in the hierarchical classification we only considered the values on the sixteen variables that were used to make the preliminary classifications in Chapter 7.

The results of this second hierarchical classification are

Table 8.3.9

Comparison of Final Classifications:

Hierarchical Classification on 428 Occupations

Hierarchical Classification	Number of Occupations in Common with Those Identified in the Conceptual Approach		
	Independent	Subordinate	Secondary
Independent	44	34	31
Subordinate	7	94	60
Secondary	10	18	55
Total Identified in the Three Analyses of the Conceptual Approach	61	146	146

illustrated by the dendogram shown in Figure 8.3.3. Just as for the dendogram in Figure 8.3.1, this dendogram only shows the groupings from 50 groups to 1 group. The most distinct group is that at the bottom of the dendogram which is labelled E and 4. Examining the data reveals that this group is composed of four farm occupations. Including this small group would distort the importance of farm occupations in the labour market. Therefore, we decided to omit this group from the initial analysis. The dotted vertical line in Figure 8.3.3 marks a level of dissimilarity at which we can obtain three relatively compact and distinct groups. These are the groups labelled 1, 2, and 3.

We have called group 1 the independent primary, group 3 the subordinate primary, and group 2 the secondary segment. The values of the means of the variables for each group are listed in Table G.1.a in Appendix G. The means suggest that the groups are sufficiently different to justify further analysis. Furthermore, the relative rankings of the means of most of the variables correspond to our expectation of the characteristics of the composition of the segments and justifies the labelling used.

A comparison of this initial classification with the preliminary classification used in the conceptual approach is summarised in Table 8.3.10. From the results shown in Table 8.3.10 we can conclude that the secondary segment group most distinctly resembles the secondary segment in the conceptual approach. The independent primary group is the most heterogeneous in terms of the composition with respect to the initial classifications in the conceptual approach.

Because we consider the independent primary segment obtained in this hierarchical classification to be more heterogeneous than we would like, we decided to reexamine the dendogram in order to obtain three relatively more homogeneous groups with respect to the composition of the classes in the conceptual approach. The smooth vertical line

Figure 8.3.3

Dendogram: Hierarchical Classification on 295 Occupations

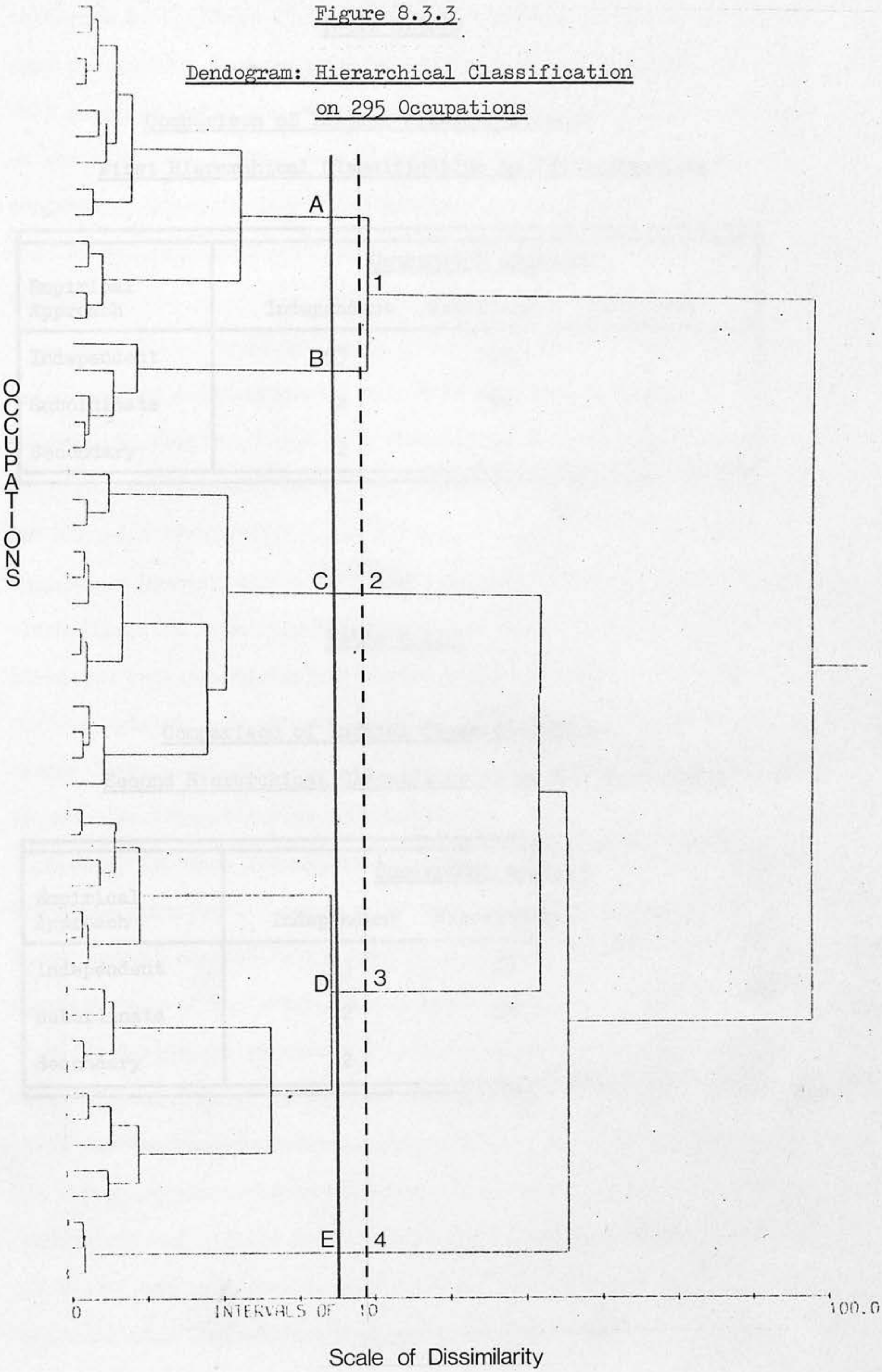


Table 8.3.10

Comparison of Initial Classifications:

First Hierarchical Classification on 295 Occupations

Empirical Approach	Conceptual Approach		
	Independent	Subordinate	Secondary
Independent	53	55	33
Subordinate	2	56	43
Secondary	2	11	40

Table 8.3.11

Comparison of Initial Classifications:

Second Hierarchical Classification on 295 Occupations

Empirical Approach	Conceptual Approach		
	Independent	Subordinate	Secondary
Independent	53	43	1
Subordinate	2	56	43
Secondary	2	11	40

in Figure 8.3.3 identifies five groups. Group E is the group of farm occupations that we already omitted from the analysis. We find that groups C and D are the same as 2 and 3, respectively. However. we now find that group 1 from the previous grouping is composed of two groups, A and B. Since the composition of B is more heterogeneous than A, we decided to omit the occupations in group B from the initial analysis.

The three new groups are A, B and D in the dendogram in Figure 8.3.3. Group A is associated with the independent primary, D with the subordinate primary, and B with the secondary segment. The values of the means and standard deviations of the variables for each segment are listed in Tables G.2.a and G.2.b in Appendix G. These values indicate a greater degree of homogeneity than appeared in the first classifications from this hierarchical classification. The relative sizes and rankings of the means across the segments correspond surprisingly well to our expectations of the conceptual framework of labour market segmentation. In addition, the comparison of this second classification with the preliminary classification in the conceptual approach summarised in Table 8.3.11 reveals that each of the three groupings is composed of more occupations from its corresponding labour market segment in the initial classifications in the conceptual approach than of the other two segments.

We decided to perform a discriminant analysis on each of these two different initial classifications in order to analyse further the sixty factors used in discriminating between the labour market segments. The three groupings for the first analysis contain 141,101, and 53 occupations and 133 are left unclassified, whilst the groupings for the second analysis contain 97,101, and 53 occupations and 177 left unclassified. The tables and figures for these two discriminant analyses are presented in Appendix G. The tables and the figure labelled G1 correspond to the analysis on the first groupings, and those labelled G.2 correspond to the analysis on the second groupings.

In the first analysis we find in Table G.1.c that fifty variables are used and that twenty-two are significant discriminators. In the second analysis we find in Table G.2.c that forty-nine variables are used and twenty-five are significant discriminators between labour market segments. In both discriminant analyses the seven most successful discriminators of the labour market structure are the same. The three most important are description as White Collar (X57), description as Blue Collar (X58), and proportion of Blacks (X4). Total personal income (X48) is not a significant disciminator in either analysis. Significance at the .01 level of the F-statistics for the F-matrices in Tables G.1.d. and G.2.d is accepted as evidence that both groupings are distinct and worth considering further.

In these discriminant analyses we use the information on all sixty variables. In the first, forty-six of the variables in the discriminant functions presented in Table G.1.e are the same as those used in the final analysis using sixty variables in the conceptual approach, and in the second forty-four of the variables in the functions listed in Table G.2.e are the same as those in the final analysis using sixty variables in the conceptual approach. The coefficients for fifteen variables and thirteen variables from the first and second analyses, respectively, allow us to draw conclusions similar to those presented in subsection 8.2.a. For the most part the relative rankings of the coefficients for the variables was out of line for only one segment with the rankings depicted in the conceptual approach. Just as in the discussion of the hierarchical classification on 428 occupations, the coefficients for some variables are at variance with our expectations, whilst others reinforce our expectations. These results, too, would be due to the degree of heterogeneity of these groups.

Surprisingly, the discriminant analyses on the two initial classifications produced the same number of misclassifications as seen in Tables G.1.f and G.2.f. The evidence on the significance of both classifications summarised in Tables G.1.g and G.2.g is nearly the same. In the first, 49.5% were core members and 17.6% were non-members. Similarly, in the second, 47.8% were core members and 15.5% were non-members. In both analyses, the classifications were less cohesive and distinct than those obtained in the conceptual approach. The two dimensional representations of these two classifications pictured in Figures G.1 and G.2 graphically illustrate that even though these groups appear to be more distinct, they are less compact internally.

The final classification for the analysis of the first classifications has 201 in the independent primary, 157 in the subordinate primary, and 70 in the secondary segment. Surprisingly, the final classification for the second analysis is 200, 157, and 71, respectively. Comparison of these final classifications with the occupations commonly identified by the three sets of analyses performed in the conceptual approach are shown in Tables 8.3.12, and 8.3.13, respectively for the first groupings and the second groupings. In both comparisons, each labour market segment has the largest number of occupations corresponding to the occupations in the respective labour market segment obtained in the conceptual approaches. Just as in the discriminant analysis of the hierarchical classification on all 428 occupations, the secondary segment is the most similar to its corresponding labour market segment in the conceptual approach, and the independent primary is the most heterogeneous.

Table 8.3.12

Comparison of Final Classifications:

First Hierarchical Classification on 295 Occupations

Hierarchical Classification	Number of Occupations in Common with Those Identified in the Conceptual Approach		
	Independent	Subordinate	Secondary
Independent	59	45	44
Subordinate	2	89	48
Secondary	0	12	54
Total Identified in the Three Analyses of the Conceptual Approach	61	146	146

Table 8.3.13

Comparison of Final Classifications:

Second Hierarchical Classification on 295 Occupations

Hierarchical Classification	Number of Occupations in Common with Those Identified in the Conceptual Approach		
	Independent	Subordinate	Secondary
Independent	59	45	44
Subordinate	2	88	49
Secondary	0	13	51
Total Identified in the Three Analyses of the Conceptual Approach	61	146	146

8.4 Summary

In the conceptual approach in which we applied the technique of discriminant analysis to the preliminary classifications described in Chapter 7, we were able to analyse the factors that determine the labour market structure and to assign occupations to the labour market segments. Our results indicate that the single most important factor affecting the classification into a labour market segment is total personal income. However, more important is that in each of the three sets of discriminant analyses in this conceptual approach over 60% of the variables used were significant discriminators of the labour market structure. In particular, in the analysis using the initial list of fifty-nine variables, fifty-one were used in the determination of the segments, and thirty-eight were significant discriminators. The more significant of the variables included education, chief income recipient, percentage of males, percentage of Whites, percentage of Blacks, age, urban residence, central city residence, employment in the central city, type of transportation to work, employment in the government sector, percentage of persons below the poverty level, and job description.

Furthermore, the technique of discriminant analysis does provide a means for explaining the concept of labour market segmentation more specifically. Occupations and, in turn, individuals in the three labour market segments can be distinguished one from another. Discriminant analysis highlighted the differences between the two segments of the primary labour market segment. In fact, the occupations in the independent primary do offer more flexibility to individuals in choosing the location of employment and working patterns. Even though there are great differences between the independent primary and secondary segments, for example, in terms of education and income, we have seen that for a number of variables, particularly those on the work climate and patterns, these two segments are characterised by similar coefficients.

We were not surprised to find that sex, race, and education are significant factors describing the characteristics of the labour market segments. The coefficients for most of these variables confirmed our expecations that men are preferred to women, Whites to Blacks, higher levels of educational attainment to lower levels. From the results we found that particular industrial sectors favoured occupations in one of the three segments. But, the coefficients for these variables were not as marked as expected. Furthermore, the coefficients for the type of employment, that is, private, government or self-employment, indicated that all three sectors of the economy maintain occupations and offer employment in similar proportions in all three labour market segments. Generally, then, the labour market segments seem to cut across all sectors of industry and the economy more than expected. We would have expected certain industries to have a large number of occupations predominantly located in only one of the three labour market segments. The coefficients for the variables on place of residence and location of employment indicate that highly urbanised and centralised areas are more likely to be characterised by all three segments of the labour market than rural areas. In addition, the occupations and workers of the secondary segment are less likely to be located outside the SMSA of the highly urbanised areas.

In the empirical approach we found that with the technique of hierarchical classification we could create groupings of occupations and individuals in the labour market. Even though the groupings did not mirror perfectly the composition of the labour market segments obtained in the conceptual approach, they did bear a surprisingly reasonable likeness in terms of the relative rankings of the coefficients for most of the variables. However, some of the results only served to add ambiguity to an explanation of the differences between the labour market segments. It is difficult to provide a complete and thorough explanation of the variance of some of the results with the concept of labour market segmentation. The unexpected results for some of the variables are undoubtedly due in large part to the heterogeneity of the composition of the groups with respect to the classifications obtained in the conceptual approach that cannot be eliminated. Nevertheless, each of the three different groupings obtained in the two hierarchical classifications did resemble its corresponding labour market segment more than the other two. We did not expect a duplication of the classifications obtained in the conceptual approach, but the predominant occupations in each grouping were also those that were identified with their respective labour market segments in the conceptual approach. The results of the empirical approach, unbiased by subjective considerations, indicate that segments in the labour market can be identified and are characterised by the factors described in Section 7.3 on the concept of labour market segmentation. The discriminant analysis of these empirical hierarchical classifications strengthened our expectations about the influences of a number of the statistically significant variables on the characteristics of the labour market structure. The heterogeneous composition of the independent primary segment can be considered an indication of the stronger associations for this segment with the other two than for any one of the other two segments with the remaining two segments.

CHAPTER 9

Concluding Remarks

9.1 Summary

The purpose of this thesis has been to make a contribution to our conceptual and empirical understanding of the structure of the labour market taking into consideration the factors that limit individual choice and are responsible for the inequality of personal income. In doing so, the scope of this thesis has been limited to explaining the relations between the groups of buyers and sellers of labour in the labour market without directly undertaking to elaborate upon the subsequent consequences on the issue of equity. The work in this thesis sought to accommodate the changing thinking of academic economists that recognises the integrated processes of production and distribution in the labour market and the importance of the labour market decisions of buyers and sellers in the functioning of these processes.

In the United States, as in other Western economies, it is recognised that the distribution of personal income is unequal, and, in addition, has not been changing markedly through the years. A number of theories and explanations have been put forward to account for the facts and the nature of the distribution of personal income. Chapter 2 reviewed these theories and explanations of the distribution of personal income. Even though we found that these theories and explanations were unsatisfactory, the review enabled us to identify a number of the important factors that underlie the distribution. Furthermore, from this discussion we noted that the existing theories and explanations recognised the importance of individual choice in the earnings and distribution processes, but did not incorporate the influence and contribution of the labour market structure on individual choice.

In fact, we found that the earnings and distribution processes occur within the labour market structure. An individual's ability to earn income is determined by his positionin the labour market. It is the labour market structure that limits an individual's possibilities for earning an income. Therefore, if we wish to determine the influence of the labour market on the earnings and distribution process, we must understand the structure and the operations ofthe labour market. Chapter 3 provided a discussion of labour market analysis. In this discussion, we found that the conventional approach could not accommodate all the real-world factors that determined the decisions of the buyers and sellers of labour. We found that differences in sex, race, education, and other non-market and non-competitive factors have a significant influence indeterminingthe treatment of an individual in the labour market, and that groups of buyers and sellers of labour participate in the labour market. This led us to examine the development of alternative approaches to the labour market. These alternative approaches attempt to develop conceptual frameworksof the labour market that recognise the social, institutional, and economic factors influencing the limits on individual choice. One such alternative is Piore's concept of labour market segmentation that analyses a labour market distinguished by three segments with different characteristics.

Having established the feasibility of the concept of labour market segmentation as an alternative framework for labour market analysis, we next turned to examine the technical and conceptual considerations involved in adopting this concept. We found the concept of labour market segmentation has been suggested in broad and general terms, but has not been subjected to rigorous empirical analysis.

It was in Chapter 4 that we undertook to put forward and discuss the problems of developing the concept of labour market segmentation into a systematic, testable, and workable framework.

The first task performed in the development of the concept of labour market segmentation was the acquisition of data that would accommodate an empirical analysis. In Chapter 5, we found that there was no data source that was specifically developed to reflect the concept of labour market segmentation. The most suitable and available data were 1970 U.S. Census data. Since we wished to form groups of occupations that would reflect the segments of the concept of labour market segmentation, we developed a procedure that transformed the 1970 U.S. Census data on individuals into data on occupations. This transformed data on occupations served as the data source for the empirical analysis in the thesis. The data source is composed of 428 occupations on which were observed sixty variables.

The second task that was necessary to perform was the selection of the most appropriate statistical techniques and procedures. We wished to group together those occupations that were most similar and to separate those occupations that were most similar and to separate those occupations most unalike. In other words, we sought to find a statistical technique that would enable us to form three groups of occupations that would represent the three labour market segments of labour market segmentation and that would reflect the differences in labour market characteristics between the segments. As discussed in Chapter 6, the techniques of classification analysis would provide suitable statistical techniques for obtaining the most homogeneous groups of occupations. We decided to use the techniques of discriminant analysis and hierarchical classification analysis for the empirical work in this thesis. Chapter 6 discussed the statistical foundations of these two techniques of analysis and the method of interpreting their results.

The third task, and probably the most important, was to elaborate upon the application of the two techniques of classification analysis to the analysis of the concept of labour market segmentation. We decided in Chapter 7 to perform two parallel approaches. The first, the conceptual approach, would be an analysis of groups of occupations organised through objective and subjective considerations to reflect the three labour market segments. The second, the empirical approach, would be an analysis of groups obtained in an empirical classification based entirely upon the statistical similarity between occupations. But, if we wished to develop any conclusions from the results of these analyses, we must develop a realistic and relevant framework from the concept of labour market segmentation. It was to this purpose that we next turned in Chapter 7. We developed a framework of sixty socio-economic variables and hypothesised about the expected influence of each variable on the labour market structure. It is this framework developed in Chapter 7 that would serve as the basic reference with which we could identify and analyse the quantitative significance of the three labour market segments.

Finally, in Chapter 7 we selected a sample of occupations and classified them into three groups on the basis of the values for sixteen variables and subjective judgement. This preliminary classification served as the starting point from which to begin the discriminant analysis in the conceptual approach. Having developed a testable and workable framework of labour market segmentation and an initial classification into the labour market segments, we next conducted the analysis as discussed in Chapter 8. In Chapter 8, we firstly considered the conceptual approach. In this approach we performed three sets of discriminant analysis. We found in all three that discriminant analysis could identify three relatively distinct groups of occupations and could facilitate the analysis of the influence of the significant variables in discrminating between the groups. Despite the varying amounts of information on which the three sets of analysis in the conceptual approach were based, the results of all three created labour market segments with quite similar composition. The results confirmed most of our expectations about the influence of the different factors in determining the labour market segments. We, then, turned in Chapter 8 to the empirical approach. It was in this empirical approach that we hoped to establish whether or not the labour market segments could be obtained without the influence of subjective judgement. To this end, we used the technique of hierarchical classification to classify statistically the occupations into three groups. Having obtained the groups, we applied the technique of discriminant analysis to analyse the groups and to make the final assignments to segments. On the basis of the results in this approach, we suggested that the concept of labour market segmentation does, in fact, provide one suitable representation of the realities of the labour market. Bearing in mind that the statistical formation of the groups depended upon the nature of the variables, we further suggested that identifiable segments do exist in the labour market and can be distinguished by such general characteristics as the barriers to mobility, the family and social structure, the geographic location of residence and employment, the industrial structure, the work climate, the economic performance, and the institutional structure.

9.2 General Conclusions

The intentions which were set forth at the beginning of this thesis were to provide a general understanding of the concept of labour market segmentation and to analyse the factors that distinguish Piore's three different labour market segments. In doing so, we attempted to establish how social, institutional, and economic factors affect individual choice in the labour market. The thesis can be viewed as one attempt to provide an alternative framework for labour market analysis. The basic method for constructing this framework was the application of classificaftion analysis to the concept of labour market segmentation.

Through the use of the techniques of classification analysis, we established what we hope is a realistic and relevant framework of labour market segmentation. These statistical techniques enabled us to analyse systematically sixty socio-economic factors that cumulatively influence the operations of the labour market. The viewpoint established by the classification analysis showed a close correspondence between the variables selected by the discriminant analysis and our general expectations about the characteristics of the labour market structure. Many of these factors that influence the labour market structure are the same as those identified in the empirical studies on the determinants of personal income presented in Chapter 2. However, the factors identified in our analysis are embodied in a workable framework of the labour market. Classification analysis identified similarities and differences between the segments of the labour market. Furthermore, with the technique of discriminant analysis we can assess the suitability of the assignment of an occupation and an individual to a labour market segment.

Nevertheless, the classifications of occupations into labour market segments is only as good as the data in the analysis. The data used in this analysis were not ideal. As we saw in Chapter 5, the data are not collected in a form that specifically accommodates the identification of the labour market segments. We had to take the data that were available and transform them into a form that could be regarded as consistent with an analysis of the concept of labour market segmentation. With the data we identified sixty socio-economic factors. Even so, many of the important factors are not measured. For example, since Piore's concept of labour market segmentation depends to some extent on the differences between the type of work rules, lines of promotion, security of employment, on-the-job training, worker evaluation of the working conditions, and others, we would have ideally wanted to include these factors in our analysis. More data on the labour market is needed before any more precise conclusions on the characteristics of the segments can be drawn. More comprehensive data may possibly permit the specification of more homogeneous segments. However, with the limited data available, we believe we have made definite progress towards establishing the segments and identifying the overall effect of the factors determining the labour market structure.

In general, the results of the work in this thesis show that a methodology based upon the techniques of classification analysis produced a more precise conceptual and empirical understanding of the structure of the labour market than was afforded in the general discussion of the concept of labour market segmentation. Assuming that our hypotheses about the framework of labour market segmentation that we developed in Chapter 7 and on which we based our analysis are correct, the work in this thesis can be of value for further work in three general areas. Firstly, to those interested in establishing the nature of the historical, political, and economic processes that may have caused the development of a labour market composed of three separate segments, the work in this thesis highlights the key features of the labour market in the 1970's. The methodology and basic framework could be extended to test the implications of the concept of labour market segmentation across time. In this way, the predominant factors and the changes in their significance in discriminating between the segments could be assessed historically. If the factors could be identified historically, then political and social development could be viewed in the context of a model incorporating a labour market structure of three segments (or any number of segments). In short, since the methodology used in this thesis highlights the main features of the labour market in a manner easily understood, the extension of the methodology across time could possibly uncover the main features of the labour market in the process of economic development.

Secondly, the work in this thesis provides further opportunities to examine in detail the statistical nature of the interrelationships among the more important factors. In the work in this thesis, the factors were all considered to be exogenously and independently determined. However, it is very likely that all the factors are not independently determined. Because some of the results of the analysis were ambiguous and dubious, further statistical testing on the independence between variables and the implications of the results of this testing on the meaning of our quantitative results are required before more than general conclusions can be drawn.

Finally, to those interested in adopting the concept of labour market segmentation into more general economic analysis, the work in this thesis can be of value. For one, if data were available, the work could be extended to follow individuals across time through the process of deciding upon education, training, and occupations in view of the structure of the labour market across time. For another, the respective characteristics of the labour market segments can be examined to facilitate the incorporation of the labour market structure into a more comprehensive and succssful explanation of the distribution of personal income. Finally, since the work in this thesis identifies what may be considered the dominant groups in the economic system, a study of the respective roles of these groups in terms of investment and savings could help in the assessment of the theories of economic growth. With such a study, the effects of altering the labour market structure, that is, reducing the differences between segments, on the overall level of savings and investment, on the incidence of taxation, and on the incidence of poverty could possibly be revealed. Unfortunately, each of these suggestions would embrace a large, complex, and detailed study.

The main purpose of this thesis has been to present an alternative framework for labour market analysis. On the whole, the use of classification analysis provided the means for establishing the labour market segments and examining the relationships within and among the segments in a way previously untried. It is the techniques of classification analysis that enabled us to sharpen and to make more concrete the concept of labour market segmentation. This alternative framework, labour market segmentation, must be judged on its ability to explain in more detail than the conventional approach the observed phenomena that individuals of seemingly equal economic potential earn different incomes and often occupy different relative positions in the labour market. Furthermore, it must be judged on its ability to explain the operations of the labour market in terms of the interrelations between the decisions of groups of buyers and sellers.

Appendix A[1)]

1970 CENSUS OF POPULATION--OCCUPATIONAL CLASSIFICATION SYSTEM

"n.e.c." means not elsewhere included.
Asterisks identify occupational classifications not in the public use sample.

Occupation Code	PROFESSIONAL, TECHNICAL, AND KINDRED WORKERS
001	Accountants
002	Architects
	Computer specialists
003	Computer programmers
004	Computer systems analysts
005	Computer specialists, n.e.c.
	Engineers
006	Aeronautical and astronautical engineers
010	Chemical engineers
011	Civil engineers
012	Electrical and electronic engineers
013	Industrial engineers
014	Mechanical engineers
015	Metallurgical and materials engineers
020	Mining engineers
021	Petroleum engineers
022	Sales engineers
023	Engineers, n.e.c.
024	Farm management advisors
025	Foresters and conservationists
026 *	Home management advisors
	Lawyers and judges
030	Judges
031	Lawyers
	Librarians, archivists, and curators
032	Librarians
033	Archivists and curators
	Mathematical specialists
034	Actuaries
035	Mathematicians
036	Statisticians
	Life and physical scientists
042	Agricultural scientists
043	Atmospheric and space scientists
044	Biological scientists
045	Chemists
051	Geologists
052	Marine scientists
053	Physicists and astronomers
054 *	Life and physical scientists, n.e.c.
055	Operations and systems researchers and analysts
056	Personnel and labor relations workers
	Physicians, dentists, and related practitioners
061	Chiropractors
062	Dentists
063	Optometrists
064	Pharmacists
065	Physicians, medical and osteopathic
071	Podiatrists
072	Veterinarians
073	Health practitioners, n.e.c.

Occupation Code	PROFESSIONAL, TECHNICAL, AND KINDRED WORKERS—Continued
	Nurses, dietitians, and therapists
074	Dietitians
075	Registered nurses
076	Therapists
	Health technologists and technicians
080	Clinical laboratory technologists and technicians
081	Dental hygienists
082	Health record technologists and technicians
083	Radiologic technologists and technicians
084	Therapy assistants
085	Health technologists and technicians, n.e.c.
	Religious workers
086	Clergymen
090	Religious workers, n.e.c.
	Social scientists
091	Economists
092*	Political scientists
093	Psychologists
094	Sociologists
095	Urban and regional planners
096	Social scientists, n.e.c.
	Social and recreation workers
100	Social workers
101	Recreation workers
	Teachers, college and university
102	Agriculture teachers
103	Atmospheric, earth, marine, and space teachers
104	Biology teachers
105	Chemistry teachers
110	Physics teachers
111	Engineering teachers
112	Mathematics teachers
113	Health specialties teachers
114	Psychology teachers
115	Business and commerce teachers
116	Economics teachers
120	History teachers
121	Sociology teachers
122	Social science teachers, n.e.c.
123	Art, drama, and music teachers
124	Coaches and physical education teachers
125	Education teachers
126	English teachers
130	Foreign language teachers
131	Home economics teachers
132	Law teachers
133	Theology teachers
134	Trade, industrial, and technical teachers
135	Miscellaneous teachers, college and university
140	Teachers, college and university, subject not specified

1) U.S. Department of Commerce(1972a, pp. 6-10).

Occupation Code	PROFESSIONAL, TECHNICAL, AND KINDRED WORKERS—Continued
	Teachers, except college and university
141	Adult education teachers
142	Elementary school teachers
143	Prekindergarten and kindergarten teachers
144	Secondary school teachers
145	Teachers, except college and university, n.e.c.
	Engineering and science technicians
150	Agriculture and biological technicians, except health
151	Chemical technicians
152	Draftsmen
153	Electrical and electronic engineering technicians
154	Industrial engineering technicians
155	Mechanical engineering technicians
156*	Mathematical technicians
161	Surveyors
162	Engineering and science technicians, n.e.c.
	Technicians, except health, and engineering and science
163	Airplane pilots
164	Air traffic controllers
165	Embalmers
170	Flight engineers
171	Radio operators
172	Tool programmers, numerical control
173	Technicians, n.e.c.
174	Vocational and educational counselors
	Writers, artists, and entertainers
175	Actors
180	Athletes and kindred workers
181	Authors
182	Dancers
183	Designers
184	Editors and reporters
185	Musicians and composers
190	Painters and sculptors
191	Photographers
192	Public relations men and publicity writers
193	Radio and television announcers
194	Writers, artists, and entertainers, n.e.c.
195	Research workers, not specified

MANAGERS AND ADMINISTRATORS, EXCEPT FARM

Occupation Code	
201	Assessors, controllers, and treasurers; local public administration
202	Bank officers and financial managers
203	Buyers and shippers, farm products
205	Buyers, wholesale and retail trade
210	Credit men
211	Funeral directors
212	Health administrators
213	Construction inspectors, public administration
215	Inspectors, except construction, public administration
216	Managers and superintendents, building
220	Office managers, n.e.c.
221	Officers, pilots, and pursers; ship
222	Officials and administrators; public administration, n.e.c.
223	Officials of lodges, societies, and unions
224	Postmasters and mail superintendents

Occupation Code	MANAGERS AND ADMINISTRATORS, EXCEPT FARM—Continued
225	Purchasing agents and buyers, n.e.c.
226	Railroad conductors
230	Restaurant, cafeteria, and bar managers
231	Sales managers and department heads, retail trade
233	Sales managers, except retail trade
235	School administrators, college
240	School administrators, elementary and secondary
245	Managers and administrators, n.e.c.

SALES WORKERS

Occupation Code	
260	Advertising agents and salesmen
261	Auctioneers
262	Demonstrators
264	Hucksters and peddlers
265	Insurance agents, brokers, and underwriters
266	Newsboys
270	Real estate agents and brokers
271	Stock and bond salesmen
281	Sales representatives, manufacturing industries (Ind. 107-399)
282	Sales representatives, wholesale trade (Ind. 017-058, 507-599)
283	Sales clerks, retail trade (Ind. 608-699 except 618, 639, 649, 667, 668, 688)
284	Salesmen, retail trade (Ind. 607, 618, 639, 649, 667, 668, 688)
285	Salesmen of services and construction (Ind. 067-078, 407-499, 707-947)

CLERICAL AND KINDRED WORKERS

Occupation Code	
301	Bank tellers
303	Billing clerks
305	Bookkeepers
310	Cashiers
311	Clerical assistants, social welfare
312	Clerical supervisors, n.e.c.
313	Collectors, bill and account
314	Counter clerks, except food
315	Dispatchers and starters, vehicle
320	Enumerators and interviewers
321	Estimators and investigators, n.e.c.
323	Expediters and production controllers
325	File clerks
326	Insurance adjusters, examiners, and investigators
330	Library attendants and assistants
331	Mail carriers, post office
332	Mail handlers, except post office
333	Messengers and office boys
334	Meter readers, utilities
	Office machine operators
341	Bookkeeping and billing machine operators
342	Calculating machine operators
343	Computer and peripheral equipment operators
344	Duplicating machine operators

Occupation Code	CLERICAL AND KINDRED WORKERS—Continued
	Office machine operators—Continued
345	Key punch operators
350	Tabulating machine operators
355	Office machine operators, n.e.c.
360	Payroll and timekeeping clerks
361	Postal clerks
362	Proofreaders
363	Real estate appraisers
364	Receptionists
	Secretaries
370	Secretaries, legal
371	Secretaries, medical
372	Secretaries, n.e.c.
374	Shipping and receiving clerks
375	Statistical clerks
376	Stenographers
381	Stock clerks and storekeepers
382	Teacher aides, exc. school monitors
383*	Telegraph messengers
384	Telegraph operators
385	Telephone operators
390	Ticket, station, and express agents
391	Typists
392	Weighers
394	Miscellaneous clerical workers
395	Not specified clerical workers

CRAFTSMEN AND KINDRED WORKERS

Occupation Code	CRAFTSMEN AND KINDRED WORKERS
401	Automobile accessories installers
402	Bakers
403	Blacksmiths
404	Boilermakers
405	Bookbinders
410	Brickmasons and stonemasons
411	Brickmasons and stonemasons, apprentices
412	Bulldozer operators
413	Cabinetmakers
415	Carpenters
416	Carpenter apprentices
420	Carpet installers
421	Cement and concrete finishers
422	Compositors and typesetters
423	Printing trades apprentices, exc. pressmen
424	Cranemen, derrickmen, and hoistmen
425	Decorators and window dressers
426	Dental laboratory technicians
430	Electricians
431	Electrician apprentices
433	Electric power linemen and cablemen
434	Electrotypers and stereotypers
435	Engravers, exc. photoengravers
436	Excavating, grading, and road machine operators; exc. bulldozer
440	Floor layers, exc. tile setters
441	Foremen, n.e.c.
442	Forgemen and hammermen
443	Furniture and wood finishers
444	Furriers
445	Glaziers
446	Heat treaters, annealers, and temperers
450	Inspectors, scalers, and graders; log and lumber
452	Inspectors, n.e.c.

Occupation Code	CRAFTSMEN AND KINDRED WORKERS—Continued
453	Jewelers and watchmakers
454	Job and die setters, metal
455	Locomotive engineers
456	Locomotive firemen
461	Machinists
462	Machinist apprentices
	Mechanics and repairmen
470	Air conditioning, heating, and refrigeration
471	Aircraft
472	Automobile body repairmen
473	Automobile mechanics
474	Automobile mechanic apprentices
475	Data processing machine repairmen
480	Farm implement
481	Heavy equipment mechanics, incl. diesel
482	Household appliance and accessory installers and mechanics
483	Loom fixers
484	Office machine
485	Radio and television
486	Railroad and car shop
491*	Mechanic, exc. auto, apprentices
492	Miscellaneous mechanics and repairmen
495	Not specified mechanics and repairmen
501*	Millers; grain, flour, and feed
502	Millwrights
503	Molders, metal
504*	Molder apprentices
505	Motion picture projectionists
506	Opticians, and lens grinders and polishers
510	Painters, construction and maintenance
511*	Painter apprentices
512	Paperhangers
514	Pattern and model makers, exc. paper
515	Photoengravers and lithographers
516	Piano and organ tuners and repairmen
520	Plasterers
521*	Plasterer apprentices
522	Plumbers and pipe fitters
523	Plumber and pipe fitter apprentices
525	Power station operators
530	Pressmen and plate printers, printing
531	Pressman apprentices
533	Rollers and finishers, metal
534	Roofers and slaters
535	Sheetmetal workers and tinsmiths
536	Sheetmetal apprentices
540	Shipfitters
542	Shoe repairmen
543	Sign painters and letterers
545	Stationary engineers
546	Stone cutters and stone carvers
550	Structural metal craftsmen
551	Tailors
552	Telephone installers and repairmen
554	Telephone linemen and splicers
560	Tile setters
561	Tool and die makers
562	Tool and die maker apprentices
563	Upholsterers
571	Specified craft apprentices, n.e.c.
572	Not specified apprentices

Occupation Code	CRAFTSMEN AND KINDRED WORKERS—Continued
575	Craftsmen and kindred workers, n.e.c.
580	Former members of the Armed Forces

OPERATIVES, EXCEPT TRANSPORT

601 Asbestos and insulation workers
602 Assemblers
603 Blasters and powdermen
604 Bottling and canning operatives
605 Chainmen, rodmen, and axmen; surveying
610 Checkers, examiners, and inspectors; manufacturing
611 Clothing ironers and pressers
612 Cutting operatives, n.e.c.
613 Dressmakers and seamstresses, except factory
614 Drillers, earth
615 Dry wall installers and lathers
620 Dyers
621 Filers, polishers, sanders, and buffers
622 Furnacemen, smeltermen, and pourers
623 Garage workers and gas station attendants
624 Graders and sorters, manufacturing
625 Produce graders and packers, except factory and farm
626 Heaters, metal
630 Laundry and dry cleaning operatives, n.e.c.
631 Meat cutters and butchers, exc. manufacturing
633 Meat cutters and butchers, manufacturing
634 Meat wrappers, retail trade
635 Metal platers
636 Milliners
640 Mine operatives, n.e.c.
641 Mixing operatives
642 Oilers and greasers, exc. auto
643 Packers and wrappers, except meat and produce
644 Painters, manufactured articles
645 Photographic process workers
Precision machine operatives
650 Drill press operatives
651 Grinding machine operatives
652 Lathe and milling machine operatives
653 Precision machine operatives, n.e.c.
656 Punch and stamping press operatives
660 Riveters and fasteners
661 Sailors and deckhands
662 Sawyers
663 Sewers and stitchers
664 Shoemaking machine operatives
665 Solderers
666 Stationary firemen
Textile operatives
670 Carding, lapping, and combing operatives
671 Knitters, loopers, and toppers
672 Spinners, twisters, and winders
673 Weavers
674 Textile operatives, n.e.c.
680 Welders and flame-cutters
681 Winding operatives, n.e.c.
690 Machine operatives, miscellaneous specified
692 Machine operatives, not specified
694 Miscellaneous operatives
695 Not specified operatives

Occupation Code	TRANSPORT EQUIPMENT OPERATIVES
701	Boatmen and canalmen
703	Bus drivers
704	Conductors and motormen, urban rail transit
705	Deliverymen and routemen
706	Fork lift and tow motor operatives
710*	Motormen; mine, factory, logging camp, etc.
711	Parking attendants
712	Railroad brakemen
713	Railroad switchmen
714	Taxicab drivers and chauffeurs
715	Truck drivers

LABORERS, EXCEPT FARM

740 Animal caretakers, exc. farm
750 Carpenters' helpers
751 Construction laborers, exc. carpenters' helpers
752 Fishermen and oystermen
753 Freight and material handlers
754 Garbage collectors
755 Gardeners and groundskeepers, exc. farm
760 Longshoremen and stevedores
761 Lumbermen, raftsmen, and woodchoppers
762 Stock handlers
763 Teamsters
764 Vehicle washers and equipment cleaners
770 Warehousemen, n.e.c.
780 Miscellaneous laborers
785 Not specified laborers

FARMERS AND FARM MANAGERS

801 Farmers (owners and tenants)
802 Farm managers

FARM LABORERS AND FARM FOREMEN

821 Farm foremen
822 Farm laborers, wage workers
823 Farm laborers, unpaid family workers
824* Farm service laborers, self-employed

SERVICE WORKERS, EXC. PRIVATE HOUSEHOLD

Cleaning service workers
901 Chambermaids and maids, except private household
902 Cleaners and charwomen
903 Janitors and sextons

Food service workers
910 Bartenders
911 Busboys
912 Cooks, except private household
913 Dishwashers
914 Food counter and fountain workers
915 Waiters
916 Food service workers, n.e.c., except private household

Occupation Code	SERVICE WORKERS, EXC. PRIVATE HOUSEHOLD—Continued
	Health service workers
921	Dental assistants
922	Health aides, exc. nursing
923	Health trainees
924	Lay midwives
925	Nursing aides, orderlies, and attendants
926	Practical nurses
	Personal service workers
931	Airline stewardesses
932	Attendants, recreation and amusement
933	Attendants, personal service, n.e.c.
934	Baggage porters and bellhops
935	Barbers
940	Boarding and lodging house keepers
941	Bootblacks
942	Child care workers, exc. private household
943	Elevator operators
944	Hairdressers and cosmetologists
945 *	Personal service apprentices
950	Housekeepers, exc. private household
952	School monitors
953	Ushers, recreation and amusement
954	Welfare service aides
	Protective service workers
960	Crossing guards and bridge tenders
961	Firemen, fire protection
962	Guards and watchmen
963	Marshals and constables
964	Policemen and detectives
965	Sheriffs and bailiffs

Occupation Code	PRIVATE HOUSEHOLD WORKERS
980	Child care workers, private household
981	Cooks, private household
982	Housekeepers, private household
983	Laundresses, private household
984	Maids and servants, private household

	ALLOCATION CATEGORIES[1]
196	Professional, technical, and kindred workers—allocated
246	Managers and administrators, except farm—allocated
296	Sales workers—allocated
396	Clerical and kindred workers—allocated
586	Craftsmen and kindred workers—allocated
696	Operatives, except transport—allocated
726	Transport equipment operatives—allocated
796	Laborers, except farm—allocated
806	Farmers and farm managers—allocated
846	Farm laborers and farm foremen—allocated
976	Service workers, exc. private household—allocated
986	Private household workers—allocated
991	Unemployed persons who last worked 1959 or earlier
995 *	Not allocated

[1] Those returns from the Population Census which do not have an occupation entry are allocated among the major occupation groups during computer processing. These cases are labeled with the code for the "allocation" category to which they are assigned.

Appendix B[1)]

```
*****
TITLE
*****
   HIERARCHIC FUSION USING 8 'COMBINATORIAL'
TRANSFORMATIONS OF THE SIMILARITY MATRIX.

*******
METHOD
*******
 WE START WITH N CLUSTERS, EACH CONTAINING A SINGLE
INDIVIDUAL, WHICH ARE NUMBEREC ACCORDING TO THE INPUT
ORDER OF THE INDIVIDUALS. IN EACH OF (N-1) FUSION STEPS,
THOSE TWO CLUSTERS WHICH ARE MOST 'SIMILAR' ARE COMBINED
AND THE RESULTING UNION CLUSTER IS LABELLED WITH THE
LESSER OF THE TWO CODES OF ITS CONSTITUENT CLUSTERS. IT
HAS BEEN SUGGESTED THAT THE PROCESS CAN BE STOPPED WHEN
A SIGNIFICANT DROP OR DISCONTINUITY IN THE FUSION
COEFFICIENT VALUE IS OBSERVED. WITH THIS PROGRAM, SUCH
SELECTION IS LEFT TO THE USER - HIERAR COMPLETES ALL THE
(N-1) FUSIONS AND SUMMARIZES THE SEQUENCE IN A
'DENDROGRAM TABLE' FRCM WHICH THE ASSOCIATED DENDROGRAM
CAN BE DRAWN EASILY BY HAND. THE PROGRAM ALSO PUNCHES A
DECK OF CARDS WHICH CAN BE USED AS INPUT TO PLINK WHICH
WILL DRAW THE DENDROGRAM ON THE GRAPH PLOTTER (IT MAY
BE NECESSARY TO ACTIVATE THE CARD PUNCH WITH A CONTROL
STATEMENT IN ORDER TO PRODUCE THIS DECK - SEE JOB
CONTROL SPECIFICATIONS).

*********************
METHOD CF CCMPUTATION
*********************
 HIERAR ASSUMES THAT A SIMILARITY MATRIX HAS BEEN
COMPUTEC WITH PROGRAM CCRREL, AND PRODUCES THE FUSION
HIERARCHY BY MEANS OF A VARIABLE PARAMETRIC
TRANSFORMATION OF THE SIMILARITY COEFFICIENTS. THIS
TRANSFORMATION HAS BEEN THE SUBJECT OF SEVERAL
PUBLICATIONS (REFERENCES 51 22 29 36) AND IS EXPRESSED
AS FOLLOWS - LET CLUSTERS P AND Q BE FUSED, THEN THE
SIMILARITY S(R,P+Q) BETWEEN ANY CLUSTER R AND THE NEW
CLUSTER (P+Q) IS OBTAINED FROM THE FORMULA
  S(R,P+Q)=AP*S(R,P) + AQ*S(R,Q) + B*S(P,Q) +
          G*ABS( S(R,P)-S(R,Q) )
WHERE AP,AQ,B AND G ARE ASSIGNEC THE FOLLOWING VALUES -
 (1) SINGLE LINKAGE (NEAREST NEIGHBOUR)
           AP=AQ= 0.5, B=0, G=-0.5(DISSIMILARITY) OR
           G=0.5(SIMILARITY)
 (2) COMPLETE LINKAGE(FURTHEST NEIGHBOUR)
            AP=AQ= 0.5, B=0, G=0.5(DISSIMILARITY) OR
            G=-0.5(SIMILARITY)
 (3) AVERAGE LINKAGE(GROUP AVERAGE, UNWEIGHTED PAIR-
     GROUP)
      AP=NP/(NP+NQ), AQ=NQ/(NP+NQ), B=G=0
*(4) CENTROID (WEIGHTED-GRCUP)
      AP=NP/(NP+NQ), AQ=NQ/(NP+NQ), B=-AP*AQ, G=0
```

1) Wishart(1969b, pp. 37-42).

```
*(5) MEDIAN (GOWER'S METHOD)
     AP=AQ=0.5,B=-0.25, G=0
*(6) WARD'S METHOD (ERROR SUM OF SQUARES)
     AP=(NR+NP)/(NR+NP+NQ), AQ=(NR+NQ)/(NR+NP+NQ),
     B=-NR/(NR+NP+NQ), G=0
*(7) LANCE-WILLIAMS FLEXIBLE BETA METHOD
     AP=AQ= (1-BETA)/2, B=BETA, G=0
 (8) MCQUITTY'S SIMILARITY ANALYSIS
     AP=AQ=0.5, B=G=0
```

(NR,NP,NQ ARE CLUSTER SIZES, AND BETA IS A VARIABLE INPUT PARAMETER SPECIFIED BY THE USER FOR OPTION 7 ONLY)

*IMPORTANT - THE MEDIAN, WARD'S METHOD AND FLEXIBLE OPTIONS ARE ONLY MEANINGFUL OR DEFINED WHEN DISTANCE COEFFICIENTS (ICOEF=1 OR 2) HAVE BEEN COMPUTED WITH CORREL. THE CENTROID OPTION IS ONLY MEANINGFUL WHEN USED WITH DISTANCE, SIZE DIFFERENCE OR SHAPE DIFFERENCE COEFFICIENTS. HOWEVER, THESE 4 OPTIONS CAN BE USED WITH ANY OTHER SIMILARITY COEFFICIENT PROVIDED THAT IT IS UNDERSTOOD THAT THE ORIGINAL PUBLISHED METHOD IS NOT EVALUATED. FOR EXAMPLE, IF 'CENTROID' IS USED WITH CORRELATION COEFFICIENTS, THE RESULT IS NOT 'CENTROID SORTING' USING THE CORRELATION CRITERION BUT AN ARBITRARY TRANSFORMATION OF THE SIMILARITY MATRIX (SEE REFERENCE 22). PROGRAM CENTRO MUST BE USED TO COMPUTE CENTROID SORTING WITH THE CORRELATION CRITERION.

****
CHOICE OF SIMILARITY COEFFICIENT OPTION
****

THE FOLLOWING TABLE SUMMARIZES ALL POSSIBLE COMBINATIONS OF SIMILARITY COEFFICIENTS AND HIERAR OPTIONS WHICH ARE CONSISTENT WITH THE ORIGINAL PUBLISHED METHODS -

CODE	METHOD	USUAL COEFFICIENTS
1	NEAREST NEIGHBOUR (SINGLE LINKAGE)	ALL
2	FURTHEST NEIGHBOUR (COMPLETE LINKAGE)	ALL
3	GROUP AVERAGE (AVERAGE LINKAGE)	ALL
4	CENTROID	DISTANCE SIZE DIFF SHAPE DIFF
5	MEDIAN (GOWER'S METHOD)	DISTANCE
6	WARD'S METHOD (ERROR SUM)	DISTANCE
7	LANCE-WILLIAMS FLEXIBLE BETA (SUGGEST BETA=-0.25)	DISTANCE
8	MCQUITTY'S SIMILARITY ANALYSIS	ALL

THIS INFORMATION IS REPEATED WITHIN THE SIMILARITY COEFFICIENT CHART, WHERE VALID COMBINATIONS OF ICOEF (SIMILARITY COEFFICIENT CODE FOR CORREL) AND ITRAN (HIERAR METHOD OPTION) ARE INDICATED BY THE WORD 'YES'.

FOR ANY INVALID COMBINATION, HIERAR WILL PRINT AN ERROR WARNING MESSAGE BUT PROCEED WITH THE PROCESSING OF THE APPROPRIATE COEFFICIENT TRANSFORMATION USING THE STORED SIMILARITY MATRIX. SUCH USE OF HIERAR IS INTENDED AS A RESEARCH FACILITY FOR THOSE WHO HAVE A THOROUGH KNOWLEDGE OF THE COMBINATORIAL TRANSFORMATION TECHNIQUES.

****
LINKAGE METHODS. OPTIONS 1-3
****

THE THREE LINKAGE PROCEDURES CAN BE USED WITH ANY SIMILARITY COEFFICIENT AND THE RESULTS ARE CONSISTENT WITH THE ORIGINAL PUBLISHED METHODS. REMEMBER THAT CERTAIN OF THE INTERCLUSTER CRITERIA (E.G. INFORMATION STATISTIC) DEGENERATE TO DISTANCE WHEN USED AS (I,J) COEFFICIENTS COMPUTED BY CORREL. CONSEQUENTLY, THE LINKAGE OPTIONS DEGENERATE TO CLUSTERING BY DISTANCE WITH THESE INTERCLUSTER CRITERIA - SEE SIMILARITY COEFFICIENTS CHART.

(1) SINGLE LINKAGE - THE SIMILARITY BETWEEN CLUSTERS P AND Q IS DEFINED AS THE HIGHEST SINGLE SIMILARITY COEFFICIENT BETWEEN TWO INDIVIDUALS, ONE FROM EACH CLUSTER. SINGLE LINKAGE WILL FIND 'STRAGGLING' CLUSTERS, AND USUALLY FAILS WITH LARGE POPULATIONS DUE TO CHAINING.

(2) COMPLETE LINKAGE - S(P,Q) IS THE SMALLEST SINGLE SIMILARITY COEFFICIENT BETWEEN TWO INDIVIDUALS, ONE FROM EACH CLUSTER. COMPLETE LINKAGE FINDS SPHERICAL CLUSTERS, BUT IS LIABLE TO PRODUCE IRREGULAR RESULTS BECAUSE THE SIMILARITY CRITERION IS DETERMINED FOR ONLY TWO INDIVIDUALS AND DOES NOT ACCOUNT FOR GROUP STRUCTURE.

(3) AVERAGE LINKAGE - S(P,Q) IS THE AVERAGE OF ALL THE SIMILARITY COEFFICIENTS FOR PAIRS OF INDIVIDUALS, ONE FROM EACH CLUSTER. THIS OPTION IS EQUIVALENT TO THE UNWEIGHTED PAIR-GROUP METHOD OF SOKAL AND MICHENER (REFERENCE 41) AND REPRESENTS ONE OF THE EARLIEST ATTEMPTS TO TAKE ACCOUNT OF GROUP STRUCTURE. AVERAGE LINKAGE TENDS TO FIND SPHERICAL CLUSTERS, AND IS REASONABLY WELL BEHAVED - (THIS TRANSFORMATION IS ALSO PROPOSED INDEPENDENTLY BY MCQUITTY, REFERENCE 29).

****
CENTROID SORTING. OPTION 4
****

THIS OPTION PRODUCES A VALID 'CENTROID SORTING' ANALYSIS (EQUIVALENT TO THE WEIGHTED-GROUP METHOD OF SOKAL AND MICHENER - REFERENCE 41) WHEN USED WITH DISTANCE, SIZE DIFFERENCE AND SHAPE DIFFERENCE COEFFICIENTS. S(P,Q) IS THE VALUE OF THE SIMILARITY COEFFICIENT OBTAINED BY TREATING THE TWO MEAN COORDINATE VECTORS OF CLUSTERS P AND Q. THE TRANSFORMATION IS NOT MEANINGFUL WHEN USED WITH OTHER COEFFICIENTS, FOR WHICH PROGRAM CENTRO IS PROVIDED. THE RESULTS OBTAINED WITH CENTROID OFTEN EXHIBIT THE CHAINING EFFECT TO A SOMEWHAT LESSER EXTENT THAN SINGLE LINKAGE.

****
OPTIONS 5-8
****

(5) MEDIAN - ALTHOUGH GOWER APPEARS TO PROPOSE THIS TRANSFORMATION FOR ALL SIMILARITY COEFFICIENTS (REFERENCE 13), THE METHOC CAN ONLY BE INTERPRETED GEOMETRICALLY FOR DISTANCE COEFFICIENTS. THE DISTANCE S(R,P+Q) BETWEEN ANY CLUSTER R ANC THE CLUSTER WHICH RESULTS FROM THE FUSION OF P AND Q IS DEFINED AS THE DISTANCE FROM THE CENTROIC OF R TO THE MIDPOINT OF THE LINE JOINING THE CENTROIDS OF P AND Q. USERS WHO WISH TO FOLLOW GOWER'S PROPOSAL MAY SELECT MEDIAN WITH NONDISTANCE COEFFICIENTS ANC IGNORE THE NONCOMPATIBILITY WARNING MESSAGE. WITH DISTANCE, MEDIAN TENDS TO CHAIN FOR LARGE POPULATIONS.

(6) WARD'S METHOD - POSSIBLY THE BEST OF THE HIERAR OPTIONS, THIS TRANSFORMATION IS ONLY CONSISTENT WITH DISTANCE COEFFICIENTS. THE ERROR SUM OF SQUARES IS DEFINED AS THE SUM OF THE DISTANCES FROM EACH INDIVIDUAL TO THE CENTROID OF ITS PARENT CLUSTER. WARD(REFERENCE 44) PROPOSED THE HIERARCHICAL METHOD WHICH COMBINES THOSE TWO CLUSTERS P AND Q WHOSE FUSION YIELDS THE LEAST INCREASE IN THE ERROR SUM(THE SAME METHOD IS INDEPENDENTLY PROPOSED BY ORLOCI - REFERENCE 31). WISHART (REFERENCE 51) FOUND THE COMBINATORIAL TRANSFORMATION WHICH DERIVES WARD'S METHOD FROM THE DISTANCE MATRIX. THE PRINTEC COEFFICIENT VALUE S(P,Q) IS TWICE THE INCREASE IN THE ERROR SUM CAUSED BY FUSION, AND THE TOTAL ERROR SUM FOR ANY GROUPING IS OBTAINEC ON DIVISION BY TWO OF THE TOTAL SUM OF THOSE COEFFICIENT VALUES WHICH PRECEED THAT GROUPING IN THE PRINTOUT. THIS METHOD FINDS MINIMUM-VARIANCE SPHERICAL CLUSTERS.

(7) LANCE-WILLIAMS FLEXIBLE - THE SIMILARITY S(R,P+Q) BETWEEN ANY CLUSTER R AND THE UNION CLUSTER (P+Q) IS COMPUTED FROM THE VARIABLE TRANSFORMATION -

S(R,P+Q)=(S(R,P)+S(R,Q))*(1-BETA)/2+ S(P,Q)*BETA

WHERE BETA IS A VALUE SUPPLIED BY THE USER (SEE REFERENCE 22). LANCE AND WILLIAMS SUGGEST THAT THE VALUE BETA=-0.25 WILL PRODUCE RELIABLE RESULTS - WITH THIS VALUE, FLEXIBLE BEHAVES MUCH LIKE WARD'S METHOD.

(8) MCQUITTY'S SIMILARITY ANALYSIS - S(R,P+Q) IS DEFINED (REFERENCE 29) AS

(S(R,P)+S(R,Q))/2

AND CAN BE USED WITH ALL SIMILARITY COEFFICIENTS. SIMILARITY ANALYSIS CAN ALSO BE OBTAINED WITH FLEXIBLE WHEN BETA=0, AND TENCS TO CHAIN WITH LARGE POPULATIONS.

****
SIMILARITY MATRIX
****

IN A PREVIOUS DESCRIPTION OF HIERAR (REFERENCE 52), USERS HAVE BEEN WARNED THAT THE PROGRAM DESTROYS THE SIMILARITY MATRIX AT EXIT. THIS NO LONGER HAPPENS. WITH THE PRESENT VERSION OF HIERAR, THE DATA FILE IS COPIED FROM PERMANENT STORAGE (CLUSTAPE) ON TO DISK (CLUSDATA) BEFORE EXIT, THEREBY RESTORING THE ORIGINAL SIMILARITY MATRIX. AS FAR AS THE USER IS CONCERNED, HIERAR MAKES NO OVERALL CHANGE TO THE CLUSTAN DATA FILE.

Appendix C[1)]

BMD07M
STEPWISE DISCRIMINANT ANALYSIS

1. GENERAL DESCRIPTION

a. This program performs a multiple discriminant analysis in a stepwise manner. At each step one variable is entered into the set of discriminating variables. The variable entered is selected by the first of the following equivalent criteria:

(1) The variable with the largest F value (see computational procedure).
(2) The variable which when partialed on the previously entered variables has the highest multiple correlation with the groups.
(3) The variable which gives the greatest decrease in the ratio of within to total generalized variances.

A variable is deleted if its F value becomes too low. The program also computes canonical correlations and coefficients for canonical variables. It plots the first two canonical variables to give an optimal two-dimensional picture of the dispersion.

b. The output consists of:

(1) Group means and standard deviations
(2) Within groups covariance matrix
(3) Within groups correlation matrix
(4) At each step:

(a) Variables included and F to remove
(b) Variables not included and F to enter
(c) U statistic and approximate F statistic to test equality of group means
(d) Matrix of F statistics to test the equality of means between each pair of groups

(5) At certain specified steps and after the last step:

(a) Classification functions
(b) Classification matrix

1) Dixon(1971, pp. 214a-214l).

(6) For each case:

(a) The posterior probability of coming from each group
(b) Square of the Mahalanobis distance from each group

(7) Summary table. For each step of the procedure the following is tabulated:

(a) Variable entered or removed
(b) F value to enter or remove
(c) Number of variables included
(d) U statistic

(8) Eigenvalues, canonical variables and coefficients of canonical variables are printed and, optionally written on a tape. The number of canonical variables written on tape is equal to the number of original variables included in the last step.

(9) Plot of the first canonical variable against the second

(10) Residuals and canonical coefficients (optional)

c. Limitations per problem:

(1) p, number of variables $(1 \leq p \leq 80)$
(2) t, total number of groups $(2 \leq t \leq 80)$
(3) j, number of Variable Format Card(s) $(1 \leq j \leq 16)$

d. Estimation of running time and output pages per problem:

Number of seconds $= .0006\, p^2 (mp + 2n) + 60$ (for IBM 7094)

Number of pages $= .02n(m + 2k) + .01\,(pg^2 + p^2) + p + 10$

where p = number of variables
t = total number of groups
n = total number of cases
m = 1 if the canonical analysis is to be performed
0 otherwise
k = number of steps at which the cases are to be classified

2. COMPUTATIONAL PROCEDURE

Notation:

p = number of variables

g = number of groups used for the analysis. This excludes those with negative group size (see 3.d.)

t = total number of groups

n_m = number of cases in group m

n = total number of cases

x_{mki} = value of variable i for case k of group m

Assume for simplicity that the first g of the t groups are used for the analysis.

Step 1. The data are read and the following are formed:

Means $$\bar{x}_i = \frac{1}{n} \sum_{m=1}^{g} \sum_{k=1}^{n_m} x_{mki} \qquad i = 1, 2, \ldots, p$$

Group means $$\bar{x}_{mi} = \frac{1}{n_m} \sum_{k=1}^{n_m} x_{mki} \qquad \begin{array}{l} i = 1, 2, \ldots, p \\ m = 1, 2, \ldots, t \end{array}$$

Group standard deviations

$$s_{mi} = \sqrt{\frac{1}{n_m - 1} \sum_{k=1}^{n_m} (x_{mki} - \bar{x}_{mi})^2} \qquad \begin{array}{l} i = 1, 2, \ldots, p \\ m = 1, 2, \ldots, t \end{array}$$

Within and total cross-product matrices

$$W = \left\{ w_{ij} \right\} \; ; \; w_{ij} = \sum_{m=1}^{g} \sum_{k=1}^{n_m} (x_{mki} - \bar{x}_{mi})(x_{mkj} - \bar{x}_{mj})$$

$$T = \left\{ t_{ij} \right\} \; ; \; t_{ij} = \sum_{m=1}^{g} \sum_{k=1}^{n_m} (x_{mki} - \bar{x}_i)(x_{mkj} - \bar{x}_j)$$

$$i = 1, 2, \ldots, p \qquad j = 1, 2, \ldots, p$$

Within groups covariance matrix

$$V = \left\{ v_{ij} \right\} ; \; v_{ij} = \frac{1}{n-g} w_{ij} \qquad i = 1, 2, \ldots, p \quad j = 1, 2, \ldots, p$$

Within groups correlation matrix

$$R = \left\{ r_{ij} \right\} ; \; r_{ij} = \frac{w_{ij}}{\sqrt{w_{ii} \; w_{jj}}} \qquad i = 1, 2, \ldots, p \quad j = 1, 2, \ldots, p$$

Step 2. At each step of the procedure the variables are divided into two disjoint sets; those included in the discriminant functions and those not included. Assume for simplicity that the first r are included.

$$\text{Let} \quad W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$$

where W_{11} and T_{11} are $r \times r$.

$$\text{Let} \quad A = \begin{bmatrix} W_{11}^{-1} & W_{11}^{-1} W_{12} \\ W_{21} W_{11}^{-1} & W_{22} - W_{21} W_{11}^{-1} W_{12} \end{bmatrix} = \left\{ a_{ij} \right\}$$

$$\text{and} \quad B = \begin{bmatrix} T_{11}^{-1} & T_{11}^{-1} T_{12} \\ T_{21} T_{11}^{-1} & T_{22} - T_{21} T_{11}^{-1} T_{12} \end{bmatrix} = \left\{ b_{ij} \right\}$$

The following statistics are computed:

a) Coefficients and constant terms of the classification functions

$$c_{ki} = (n-g) \sum_{j=1}^{r} \overline{x}_{kj} a_{ij} \qquad \begin{array}{l} i = 1, 2, \ldots, r \\ k = 1, 2, \ldots, g \end{array}$$

$$c_{k0} = -\frac{1}{2} \sum_{i=1}^{r} c_{ki} \overline{x}_{ki} \qquad k = 1, 2, \ldots, g$$

b) The square of the Mahalanobis distance between each pair of groups

$$D_{m\ell} = \sum_{i=1}^{r} (c_{mi} - c_{\ell i})(\overline{x}_{mi} - \overline{x}_{\ell i}) \qquad m, \ell = 1, \ldots, g$$

c) The F values for testing differences between each pair of groups

$$F_{m\ell} = \frac{(n-g-r+1)n_m n_\ell}{r(n-g)(n_m+n_\ell)} D_{m\ell} \qquad m, \ell = 1, \ldots, g$$

with r and n-g-r+1 degrees of freedom.

d) F values for each variable

(1) If variable j has been entered

$$F_j = \frac{a_{jj} - b_{jj}}{b_{jj}} \quad \frac{n - r - g + 1}{g - 1}$$

with degrees of freedom g-1 and n-r-g+1

(2) If variable j has not been entered

$$F_j = \frac{b_{jj} - a_{jj}}{a_{jj}} \quad \frac{n-r-g}{g-1}$$

with degrees of freedom g-1 and n-g-r

Under the usual normality assumptions these are the likelihood ratio tests of the equality over all g groups of the conditional distribution of variable j given the (remaining) entered variables.

e) U statistic to test equality of group means

$$U = \mathrm{Det}\,(W_{11})/\mathrm{Det}\,(T_{11})$$

with degrees of freedom $(r, g-1, n-g)$

f) Approximate F statistic to test equality of group means

$$F = \frac{1 - U^{1/s}}{U^{1/s}} \cdot \frac{ms + 1 - rq/2}{rq}$$

$$\text{where } s = \sqrt{\frac{r^2 q^2 - 4}{r^2 + q^2 - 5}}, \text{ if } r^2 + q^2 \neq 5$$

$$s = 1, \text{ if } r^2 + q^2 = 5$$

$$m = n - \frac{r + q + 3}{2}$$

$$q = g - 1$$

its degrees of freedom are rq and $ms + 1 - rq/2$. If either r or q is 1 or 2, the approximation is exact.

g) Tolerance values

$$w_i = a_{ii}/t_{ii} \qquad , i = r + 1, \ldots, p$$

A variable passes the tolerance test if and only if w_i and t_i equal or exceed the value specified on the Subproblem Card.

Step 3. To move from one step to the next, one variable is added or removed from the discriminating set according to one of the following rules:

a) If there are one or more variables which are entered, have a control value of 1 and an F value less than "F to remove," the one with the smallest F will be deleted.

b) If no variable satisfies a), then from among those variables which have not been included, which pass the tolerance test, and have greatest control value, one is selected according to the rule specified on column 44 of the PROBLM Card.

If column 44 of PROBLM Card is blank: the variable selected has greatest F-to-enter;

If column 44 of PROBLM Card contains 1: the variable selected is the one which after entering minimizes

$$C_1 = \frac{1}{h_1} \sum_{\ell \neq m} \frac{1}{1 + D_{\ell m}/4}$$

when $h_1 = g(g-1)/2$. The motivation for this formula is that it tends to separate groups which are close together. Each term corresponds to an estimate of one minus the square of the multiple correlation between the classification variables and a dummy variable which identifies the corresponding pair of groups.

If column 44 of PROBLM Card contains 2: the variable selected is the one which after entering minimizes

$$C_2 = \frac{1}{h_2} \sum_{\ell \neq m} \frac{\alpha_{\ell m}}{1 + D_{\ell m}/4}$$

where $h_2 = \sum_{\ell \neq m} \alpha_{\ell m}$. The $\alpha_{\ell m}$ are specified on the Alpha Card.

This criterion is a simple generalization of C_1.

If column 44 of PROBLM Card contains 3: the variable selected is the one which, after entering, maximizes the smallest F between pairs of groups.

Step 4. When the number of variables entered is equal to one of the numbers indicated on the Subproblem Card and after the last step the following are computed for $\ell = 1, 2, \ldots, t$; $m = 1, 2, \ldots, g$; $k = 1, 2, \ldots, n_\ell$:

a) Value of the m^{th} classification function evaluated at case k of group ℓ

$$s_{\ell mk} = c_{mo} + \sum_{j=1}^{r} c_{mj} x_{\ell kj}$$

b) Posterior probability of case k in group ℓ having come from group m

$$p_{\ell mk} = \frac{\mathrm{Exp}(s_{\ell mk})}{\sum_{i=1}^{g} \mathrm{Exp}(s_{\ell il})}$$

c) Square of Mahalanobis distance of case k in group m from group ℓ

$$D^2_{\ell mk} = (n-g) \sum_{i=1}^{r} \sum_{j=1}^{r} (x_{mki} - \bar{x}_{\ell i}) a_{ij} (x_{mkj} - \bar{x}_{\ell j})$$

This may be used as a chi-square variable with r degrees of freedom for classification purposes.

Step 5. At this point let p denote the number of variables which are included after the last step and let W and T be their within and total sum of product matrices. Let B = T - W. The eigenvalue problem

$$Bu_i = \lambda_i Wu_i \qquad i = 1, 2, \ldots, p$$

is solved to find coefficients, u_i, of canonical variables and the amount of dispersion λ_i explained by each canonical variable.

The vectors are normalized so that

$$u_i' W u_j = \delta_{ij} \quad .$$

The canonical correlations $\rho_1, \rho_2, \ldots, \rho_p$ relative to the groups are then computed

$$\rho_i = (\lambda_i/(1 + \lambda_i))^{1/2} \quad .$$

For each case the first three canonical variables are computed

$$z_{mki} = \sum_{j=1}^{r} u_{ji}(x_{mkj} - \bar{x}_j) \qquad \begin{array}{l} m = 1, \ldots, g \\ k = 1, \ldots, n_m \\ i = 1, 2, 3 \end{array}$$

The first two of these are plotted on a scattergram. If called for it is stratified onto g separate plots on the basis of the value of z_{mk3}. The cutpoints used are the average of adjacent values, after ordering, of $\bar{z}_{m3} = \frac{1}{n_m} \sum_{k=1}^{n_m} z_{mk3}$.

3. REFERENCES

Anderson, T. W., Introduction to Multivariate Statistical Analysis, Wiley, 1958.

Efroymsen, M. A., "Multiple Regression Analysis," Mathematical Methods for Digital Computers, Part V, (17). Edited by A. Ralston and H. S. Wilf, Wiley, 1960.

Rao, C. R., Advanced Statistical Methods in Biometric Research, Wiley, 1962.

***

This program was written by Paul Sampson, a member of the staff of Health Sciences Computing Facility, UCLA.

Appendix D

Conceptual Approach:

Discriminant Analysis With

An Initial List of 60 Variables

Notation

a. Tables

1. INDPRI = Independent Primary Segment

2. PRISEC = Subordinate Primary Segment

3. SECSEC = Secondary Segment

b. Figures

1. I = Independent Primary Segment

2. P = Subordinate Primary Segment

3. S = Secondary Segment

4. * = Segment Means

5. ₤ or ♯ = Point of Overlap

Table D.1.a

Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE *	NUMBER OF VARIABLES INCLUDED
1	48		659.6643	1
2	47		64.5267	2
3	56		31.6191	3
4	8		30.6240	4
5	55		7.9701	5
6	50		5.6121	6
7	22		4.7050	7
8	42		3.9709	8
9	26		5.8831	9
10	41		3.3101	10
11	1		4.6846	11
12	7		3.2929	12
13	36		3.8302	13
14	39		3.2184	14
15	3		3.2903	15
16	60		2.5500	16
17	16		1.9842	17
18	32		2.0178	18
19	14		2.4818	19
20	20		1.7920	20
21	44		2.0244	21
22	23		1.4299	22
23	59		4.5416	23
24	54		1.2429	24
25	31		1.2047	25
26	34		1.2061	26
27	18		1.3345	27
28	33		1.2184	28
29	38		1.4851	29
30	57		1.1218	30
31	30		1.2903	31
32	6		0.9443	32
33	11		0.9490	33
34	52		0.9217	34
35	25		0.9300	35
36	45		1.0159	36
37	53		1.0618	37
38	37		0.8508	38
39	29		0.9362	39
40	49		0.7948	40
41	15		0.6319	41
42	5		0.5294	42
43	13		0.4638	43
44	9		0.4563	44
45	24		0.3173	45
46	51		0.2296	46
47	10		0.2850	47
48	27		0.1414	48
49	19		0.0650	49
50	43		0.0107	50

* F-ratio at .01 level with 50 and 229 degrees of freedom ≥ 1.60

Table D.1.b

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	------------	15.310	39.449
Subordinate	15.310	------------	19.064
Secondary	39.449	19.064	---------
F - Statistic**	18.332		

* F-ratio at .01 level with 50 and 229 degrees of freedom $\geq$ 1.60

** F-ratio at .01 level with 100 and 458 degrees of freedom $\geq$ 1.41

Table D.1.c

Discrminant Functions

VARIABLE	FUNCTION INDPRI	PRISUE	SECSEC
1	2.80596	2.71868	2.64939
3	18.75688	18.75462	18.46242
5	18.42813	18.10219	18.13486
6	-0.44823	-0.36781	-0.75123
7	4.68113	4.22615	4.08006
8	41.54181	39.60397	38.34637
9	0.02176	0.03559	0.02905
10	0.57914	0.54005	0.54336
11	2.59741	2.64423	2.65042
13	2.22404	2.20023	2.14638
14	2.62406	2.72769	2.78672
15	0.21450	0.25170	0.30105
16	2.49819	2.65553	2.75951
18	0.31122	0.36571	0.38286
19	1.00218	1.01860	1.01470
20	-0.55048	-0.58819	-0.64678
22	0.01531	-0.08807	-0.08026
23	-0.47134	-0.67569	-0.73593
24	2.60148	2.65764	2.65410
25	0.86495	0.82752	0.79536
26	0.33526	0.29351	0.31504
27	0.93236	0.93623	0.92473
29	0.72466	0.69994	0.67221
30	1.10395	1.01374	0.94132
31	1.62423	1.55672	1.51401
32	1.45036	1.50241	1.69029
33	0.51994	0.47113	0.47087
34	1.52171	1.49229	1.53543
36	6.13611	6.15246	6.40607
37	5.70714	5.80182	6.01322
38	6.07385	6.09000	6.34416
39	4.92998	4.82524	5.26066
41	2.59581	2.56136	2.46173
42	2.65382	2.67851	2.59707
43	6.50498	6.50996	6.49964
44	6.21504	6.24056	6.35481
45	6.25525	6.32389	6.35903
47	-3.24813	-3.10046	-3.30202
48	-0.43202	-0.70800	-0.72674
49	-0.54761	-0.60132	-0.69682
50	-1.47164	-1.30360	-1.35736
51	-1.21281	-0.96488	-1.16512
52	-6.20876	-6.69338	-6.58305
53	-16.52841	-15.18812	-15.71405
54	-2.39184	-2.23934	-2.19624
55	8.20296	8.25659	8.47081
56	3.79776	3.56837	3.75121
57	-30.59706	-32.12708	-33.87372
59	238.80157	257.77637	269.23071
60	7.57253	4.16060	3.02408
CONSTANT	-2231.45654	-2162.76123	-2108.89990

Table D.1.d

Classification Matrix

Segment Before the Analysis	Predicted Segment Independent	Subordinate	Secondary	Total
Independent	55	0	0	55
Subordinate	1	113	0	114
Secondary	0	2	110	112
Previously Unclassified	22	66	59	147
Total Classified	78	181	169	428

Table D.1.e

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	55	18	32.7	11	20.0
Subordinate	114	55	48.2	12	10.5
Secondary	112	60	53.6	22	19.6
Total	281	133	47.3	45	16.0

* Chi^2 at .10 level with 50 degrees of freedom $\leq$ 37.50

** Chi^2 at .99 level with 50 degrees of freedom $\geq$ 76.10

Figure D.1

Two-Dimensional Representation
of the Labour Market Segments

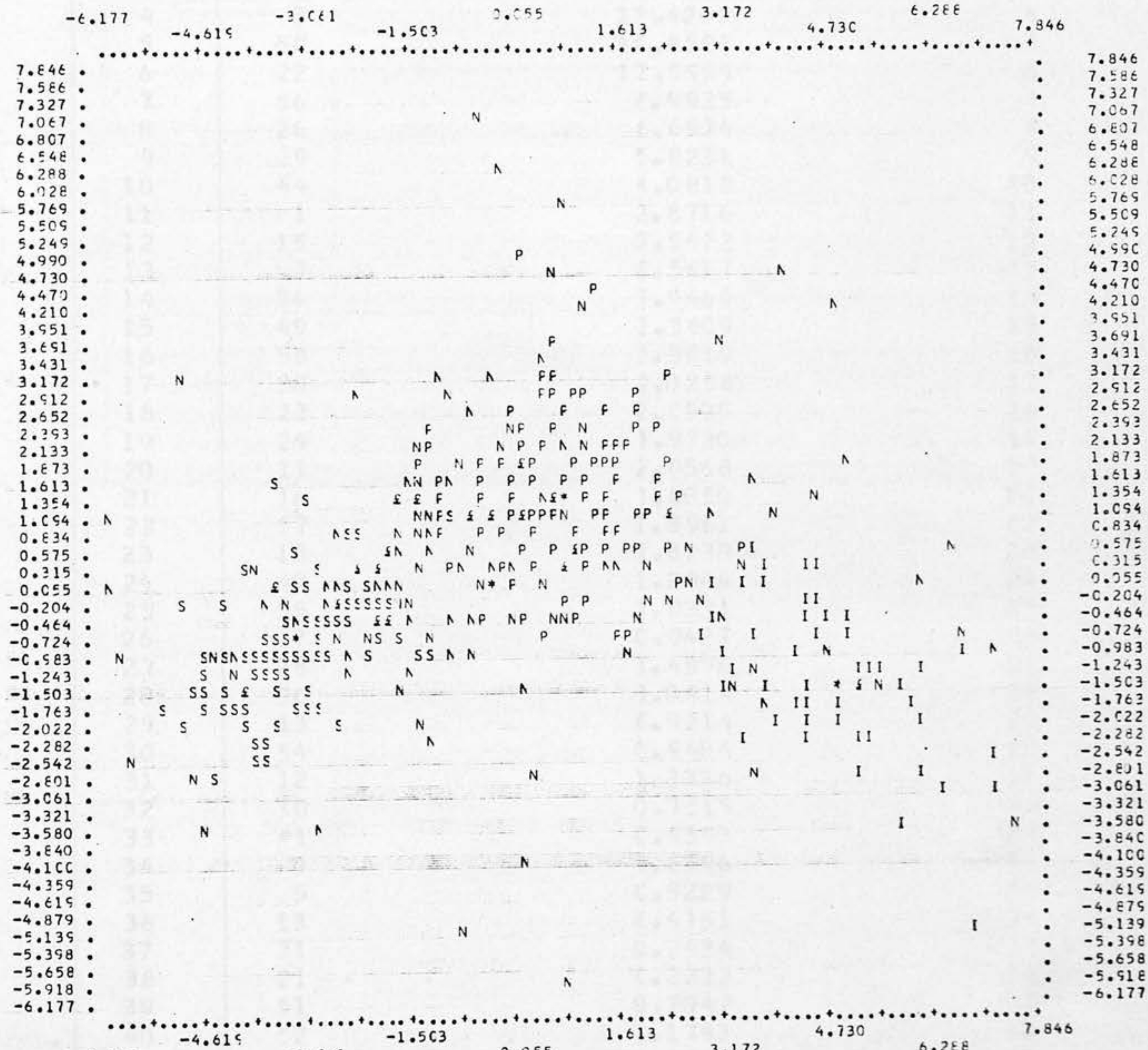


Table D.2.a

Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE *	NUMBER OF VARIABLES INCLUDED
1	48		636.2407	1
2	47		104.8274	2
3	8		32.5430	3
4	3		15.4245	4
5	50		10.8501	5
6	22		13.5959	6
7	56		6.9925	7
8	26		6.6974	8
9	39		5.8231	9
10	44		4.0818	10
11	1		2.8716	11
12	15		2.5422	12
13	7		2.5617	13
14	34		3.9464	14
15	49		3.3609	15
16	58		3.3019	16
17	38		2.0258	17
18	23		2.0535	18
19	24		1.9330	19
20	11		2.0548	20
21	16		1.8050	21
22	57		1.3961	22
23	18		1.8239	23
24	40		1.2044	24
25	55		1.0331	25
26	27		0.9427	26
27	28		1.4896	27
28	36		1.0414	28
29	13		0.9214	29
30	54		0.9404	30
31	12		1.2336	31
32	30		0.7519	32
33	41		0.9393	33
34	9		0.6596	34
35	5		0.5229	35
36	53		0.4151	36
37	31		0.2434	37
38	21		0.2232	38
39	51		0.2042	39
40	52		0.1793	40
41	14		0.1789	41
42	6		0.1508	42
43	43		0.1171	43
44	32		0.1695	44
45	25		0.0934	45
46	29		0.0475	46
47	17		0.0217	47

* F-ratio at .01 level with 47 and 237 degrees of freedom ≥ 1.61

Table D.2.b

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	----------	16.836	39.990
Subordinate	16.836	----------	22.289
Secondary	39.990	22.289	----------
F - Statistic**	20.576		

* F-ratio at .01 level with 47 and 237 degrees of freedom ≥ 1.61

** F-ratio at .01 level with 94 and 474 degrees of freedom ≥ 1.43

Table D.2.c

Discriminant Functions

VARIABLE	FUNCTION INDPRI	PRISUB	SECSEC
1	1.35532	1.27611	1.17012
3	25.35100	25.29765	24.74379
5	22.94781	22.73099	22.93869
6	5.80401	5.92921	5.79949
7	11.92919	11.50809	11.49031
8	53.90941	52.88699	51.50507
9	-0.11823	-0.10209	-0.10802
11	4.95614	5.03437	4.87056
12	1.56143	1.54252	1.58829
13	4.41621	4.36619	4.29123
14	5.46313	5.48696	5.45690
15	-3.11100	-3.05639	-2.94297
16	3.42461	3.54445	3.53786
17	1.69604	1.68545	1.67561
18	1.79339	1.85625	1.82705
21	1.10318	1.09915	1.05484
22	0.18206	0.04226	0.06853
23	-1.21478	-1.30595	-1.09450
24	15.27926	17.18645	15.81899
25	0.54984	0.53629	0.54181
26	-0.20837	-0.21212	-0.16372
27	0.30820	0.34512	0.37161
28	-0.42598	-0.39539	-0.38689
29	-0.19845	-0.19007	-0.18256
30	0.79170	0.71620	0.71487
31	0.15370	0.13249	0.12082
32	-2.88795	-2.90098	-2.82210
34	1.62448	1.69525	1.76500
36	2.14896	2.11569	2.17160
38	6.93706	6.90109	6.77222
39	-6.84157	-6.87072	-6.17438
40	2.47851	2.43415	2.52517
41	1.29802	1.22243	1.22499
43	-1.22290	-1.22538	-1.26380
44	-0.06904	-0.17337	-0.09327
47	0.93328	1.09858	0.85160
48	-1.08815	-1.46241	-1.44691
49	6.21146	6.15564	5.91930
50	-2.15148	-1.87607	-1.83369
51	-7.65510	-7.58056	-7.82505
52	-1.51171	-1.91419	-1.80724
53	2.85005	1.04145	0.28555
54	-3.21309	-3.09504	-3.20144
55	10.75905	10.78250	11.06802
56	1.58687	1.35052	1.30026
57	-26.79083	-25.18443	-27.47343
58	25.22095	29.32393	29.08423
CONSTANT	-2617.90186	-2546.41553	-2439.97412

Table D.2.d

Classification Matrix

Segment Before the Analysis	Predicted Segment Independent	Subordinate	Secondary	Total
Independent	49	1	0	50
Subordinate	0	126	0	126
Secondary	0	1	109	110
Previously Unclassified	30	58	54	142
Total Classified	79	186	163	428

Table D.2.e

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	50	16	32.0	15	30.0
Subordinate	126	57	45.2	19	15.1
Secondary	110	54	49.1	17	15.5
Total	286	127	44.4	51	17.8

* Chi^2 at .10 level with 47 degrees of freedom ≤ 34.92

** Chi^2 at .99 level with 47 degrees of freedom ≥ 73.51

Figure D.2

Two---Dimensional Representation

of the Labour Market Segments

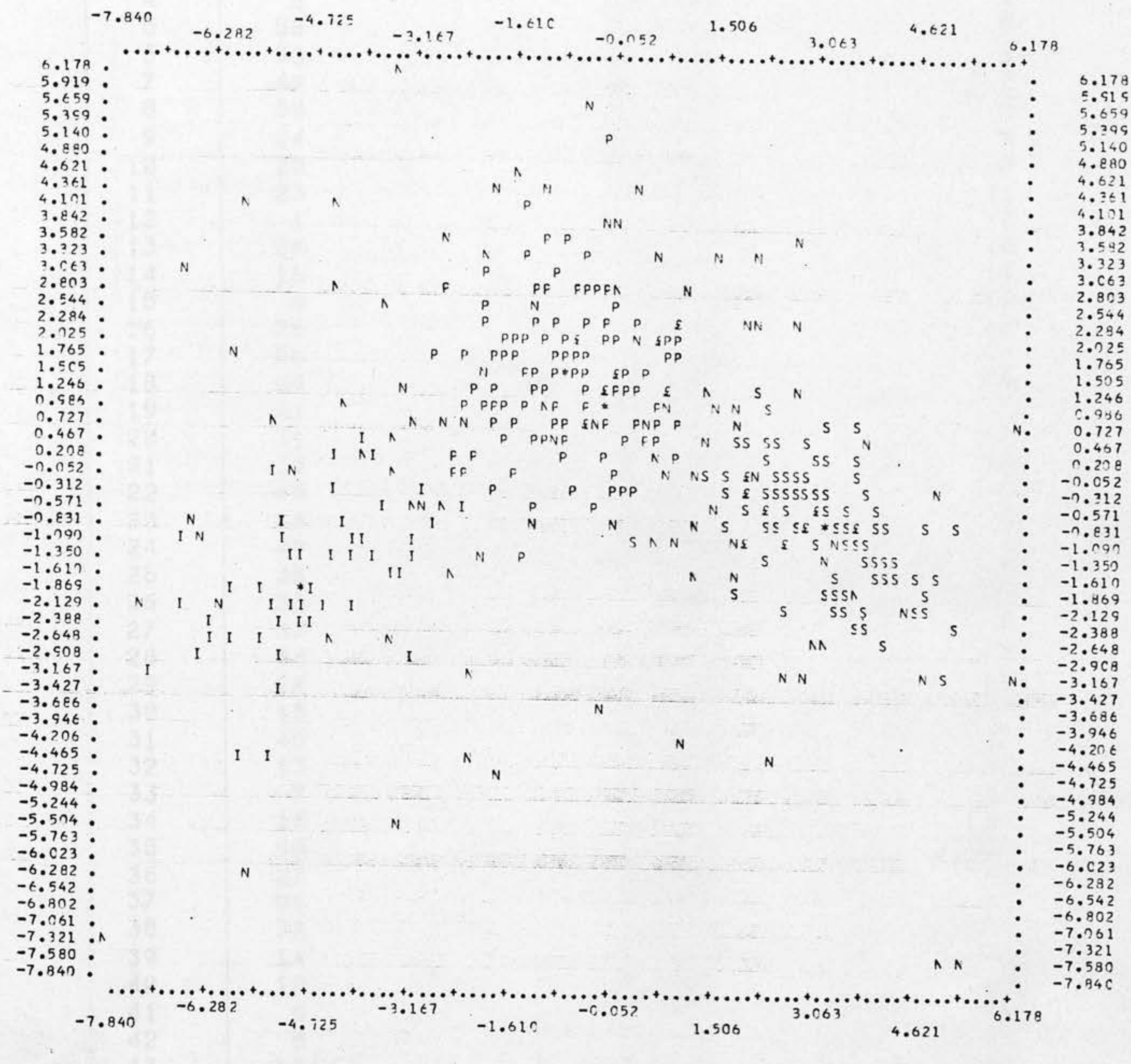


Table D.3.a
Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE *	NUMBER OF VARIABLES INCLUDED
1	48		471.3835	1
2	47		104.1029	2
3	4		33.1244	3
4	8		25.4385	4
5	56		19.9322	5
6	53		17.0242	6
7	42		12.5866	7
8	50		13.2777	8
9	24		10.8578	9
10	22		10.9550	10
11	23		7.0973	11
12	1		6.5224	12
13	58		5.5212	13
14	15		4.8098	14
15	9		4.7919	15
16	26		4.9169	16
17	55		4.3859	17
18	59		4.1583	18
19	41		2.7905	19
20	31		2.7179	20
21	39		2.3429	21
22	35		9.4696	22
23	60		2.1027	23
24	49		1.8374	24
25	36		2.1687	25
26	33		1.9380	26
27	37		1.6921	27
28	54		2.4028	28
29	20		1.7062	29
30	18		2.0413	30
31	45		1.7903	31
32	13		1.8322	32
33	7		2.1603	33
34	16		1.9943	34
35	40		1.8155	35
36	27		1.7978	36
37	51		1.1546	37
38	32		1.1708	38
39	14		1.3134	39
40	19		0.9817	40
41	6		1.1464	41
42	5		1.0045	42
43	52		0.9983	43
44	30		0.9459	44
45	12		0.9210	45
46	28		0.4988	46
47	38		0.4153	47
48	10		0.4010	48
49	44		0.3708	49
50	25		0.2023	50
51	21		0.0451	51
52	29		0.0363	52

* F-ratio at .01 level with 52 and 374 degrees of freedom ≥ 1.58

Table D.3.b

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	————	17.535	36.896
Subordinate	17.535	————	19.892
Secondary	36.896	19.892	————
F - Statistic**	20.577		

* F-ratio at .01 level with 52 and 374 degrees of freedom≥1.58

** F-ratio at .01 level with 104 and 748 degrees of freedom≥1.40

Table D.3.c

Discrminant Functions

VARIABLE	FUNCTION INDPRI	PRISUB	SECSEC
1	0.52356	0.48074	0.41497
4	-0.50967	-0.46780	-0.22327
5	-1.96612	-2.19509	-2.12574
6	-1.94535	-1.68474	-1.65527
7	4.50565	4.38718	4.39291
8	24.79716	24.22743	23.77042
9	0.12629	0.14217	0.11718
10	0.72236	0.72352	0.74680
12	1.00393	1.03013	1.07323
13	0.76283	0.73182	0.67131
14	0.42373	0.43378	0.38626
15	0.30006	0.34276	0.37878
16	0.54671	0.60264	0.63052
18	0.09618	0.17333	0.21120
19	0.39877	0.43141	0.45404
20	-1.09268	-1.17507	-1.18779
21	-1.60175	-1.61660	-1.62064
22	-0.50556	-0.62108	-0.63090
23	0.59841	0.52602	0.65966
24	0.54687	0.66604	0.64077
25	1.13932	1.14881	1.14376
26	0.79147	0.76699	0.77319
27	0.71324	0.73546	0.72593
28	0.68928	0.68416	0.67038
29	0.62602	0.61817	0.61931
30	0.32356	0.27437	0.25073
31	0.73487	0.67763	0.66152
32	0.83693	0.81999	0.74241
33	0.18995	0.16243	0.14569
35	4.14404	4.15715	3.83632
36	6.72428	6.69526	6.81971
37	6.93669	6.94303	7.05313
38	5.68315	5.66257	5.72350
39	0.43667	0.36612	0.66290
40	4.80559	4.78456	4.89045
41	4.42163	4.37181	4.43887
42	4.93981	4.99690	5.04873
44	0.56488	0.57454	0.59656
45	0.20679	0.25104	0.25059
47	-0.15690	-0.08433	-0.25528
48	-0.33512	-0.59427	-0.58476
49	-0.95433	-0.92940	-1.04557
50	-0.03316	0.15956	0.12204
51	-1.59420	-1.45769	-1.56240
52	-2.24843	-2.33611	-2.41732
53	2.94074	2.17575	1.68578
54	-0.26230	-0.23264	-0.25888
55	0.19257	0.25207	0.47998
56	-0.08782	-0.19143	-0.21515
58	27.22162	29.07704	31.03598
59	181.80624	185.00432	179.06160
60	48.34839	47.84132	50.05247
CONSTANT	-909.86182	-877.90063	-868.61963

Table D.3.d

Classification Matrix

Segment Before the Analysis	Predicted Segment Independent	Subordinate	Secondary	Total
Independent	77	2	0	79
Subordinate	1	181	4	186
Secondary	0	3	160	163
Previously Unclassified	0	0	0	0
Total Classified	78	186	163	428

Table D.3.e

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	79	36	45.6	14	17.7
Subordinate	186	119	63.9	22	11.8
Secondary	163	99	60.7	31	19.0
Total	428	254	59.3	67	15.7

* Chi^2 at .10 level with 52 degrees of freedom $\leq$ 39.25

** Chi^2 at .99 level with 52 degrees of freedom $\geq$ 78.93

Figure D.3

Two-Dimensional Representation

of the Labour Market Segments

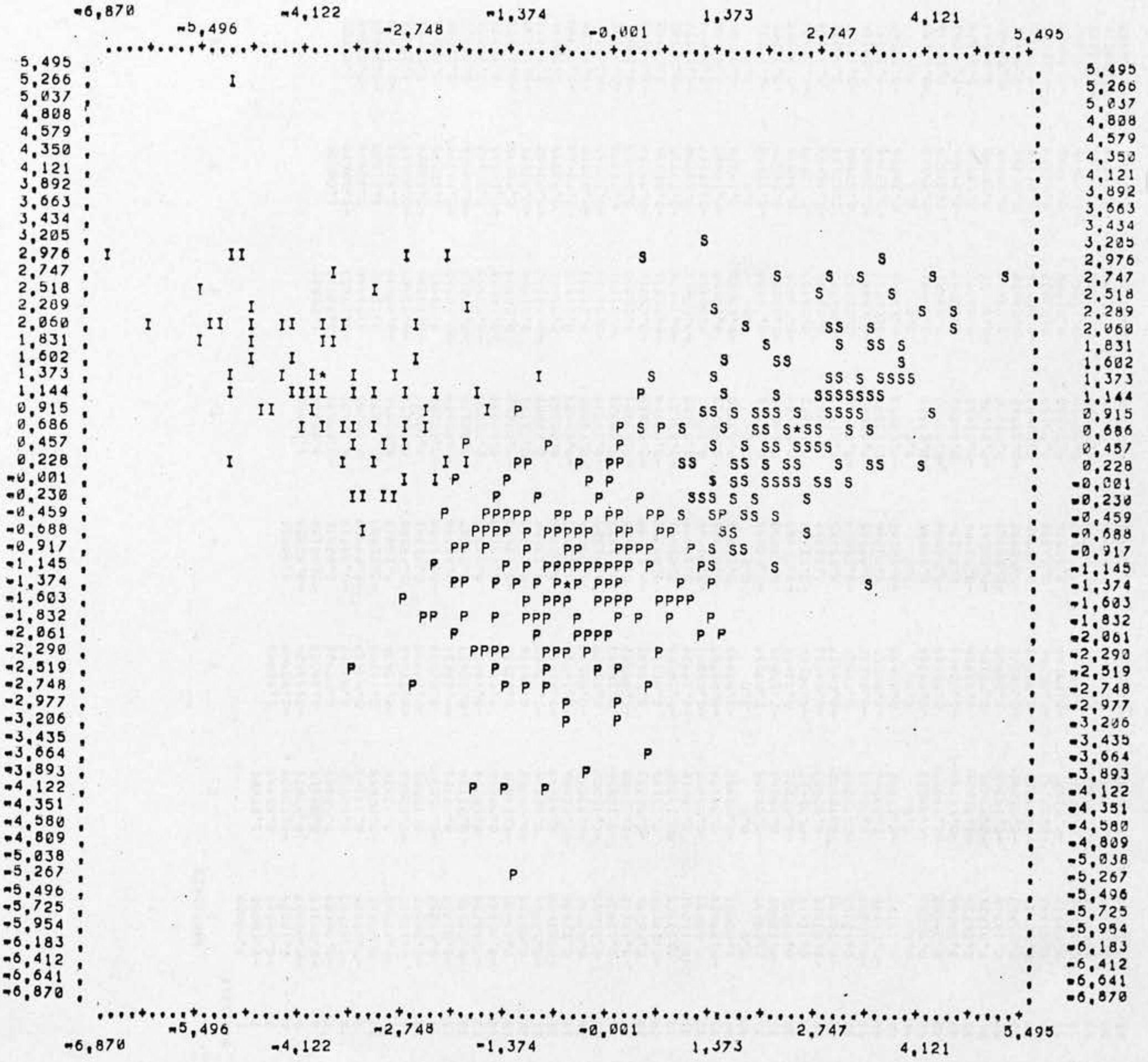


Table D.4

Correlation Matrix

VARIABLE	VARIABLES 1	2	3	4	5	6	7	8	9
1	1.00000								
2	-1.00000	1.00000							
3	-0.08847	0.08847	1.00000						
4	0.12529	-0.12529	-0.92740	1.00000					
5	-0.07877	0.07877	-0.32829	-0.03828	1.00000				
6	-0.00319	0.00319	0.04532	-0.05159	0.00874	1.00000			
7	-0.08868	0.08869	0.00926	0.02672	-0.09192	-0.02847	1.00000		
8	-0.42177	0.42176	0.02857	-0.07721	0.11671	-0.05403	-0.36428	1.00000	
9	-0.23163	0.23163	0.07437	-0.09009	0.02796	0.06446	0.21446	-0.03349	1.00000
10	0.00875	-0.00875	0.14758	-0.10527	-0.13051	0.11942	0.41620	-0.20468	0.28057
11	-0.32989	0.32989	-0.13470	0.13887	0.01051	-0.07837	0.52251	-0.11429	0.15348
12	0.15646	-0.15646	-0.07643	0.03285	0.12156	-0.07685	-0.66567	0.25601	-0.34935
13	-0.10091	0.10091	-0.20706	0.21061	0.02328	-0.13026	-0.13684	0.12005	-0.08377
14	0.05237	-0.05237	0.22978	-0.21503	-0.07285	0.05561	0.12226	-0.09216	0.04952
15	-0.01721	0.01721	-0.11224	0.12586	-0.01679	-0.07315	-0.06178	0.17638	-0.16890
16	-0.03196	0.03196	-0.32617	0.33787	0.02135	-0.02015	-0.02496	0.04415	-0.01944
17	-0.18239	0.18239	-0.17453	0.17872	0.01770	-0.00903	0.06402	0.10036	-0.01122
18	-0.00694	0.00694	0.03041	-0.03760	0.01337	-0.04956	-0.10533	0.03777	0.10706
19	0.14839	-0.14839	0.12536	-0.12420	-0.02242	0.06325	0.06038	-0.16109	-0.06332
20	0.09718	-0.09718	0.06376	-0.06251	-0.01311	0.01506	-0.02151	-0.10288	0.06519
21	-0.12864	0.12864	-0.13968	0.13684	0.02888	-0.09212	-0.02650	0.08907	-0.13212
22	-0.07658	0.07658	-0.01446	0.01382	0.00392	0.05661	0.08555	0.02835	0.02371
23	0.12606	-0.12606	0.14024	-0.12713	-0.05477	0.05979	-0.02960	-0.02788	-0.03392
24	0.03531	-0.03531	-0.08170	0.09413	-0.01851	0.21273	-0.02510	-0.07104	0.07754
25	0.31005	-0.31006	0.01400	0.01325	-0.07074	-0.04675	-0.12741	-0.20131	-0.16805
26	0.12834	-0.12834	0.11453	-0.12372	0.00528	0.17201	0.07708	-0.30585	0.15767
27	0.15152	-0.15152	-0.03902	0.05423	-0.03213	0.00077	0.04400	-0.20821	-0.14315
28	0.06251	-0.06252	0.11460	-0.11038	-0.02848	-0.05812	-0.05756	-0.02993	0.01335
29	-0.17156	0.17156	0.01173	0.00306	-0.03906	-0.06233	-0.01034	0.11313	0.01284
30	0.13710	-0.13710	-0.04159	0.06446	-0.05105	-0.10978	-0.03725	-0.02960	0.04129
31	-0.16277	0.16278	-0.18576	0.18439	0.03242	-0.00417	0.18796	-0.11757	0.02098
32	0.03893	-0.03892	-0.00541	-0.07044	0.19167	0.00854	-0.15317	0.04659	-0.13425
33	-0.39329	0.39329	-0.06793	0.03383	0.05632	0.00230	-0.12440	0.60446	-0.00775
34	0.02957	-0.02957	-0.05655	0.11773	-0.03720	-0.13921	0.13926	0.02084	0.02181
35	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
36	0.09355	-0.09355	0.07125	-0.08058	0.01242	0.15892	-0.19853	-0.25332	-0.07286
37	-0.14966	0.14966	-0.15605	0.14952	0.04066	-0.12390	0.07039	0.31289	0.00681
38	0.07222	-0.07222	0.11601	-0.08868	-0.08680	-0.05908	0.24936	-0.06557	0.11858
39	0.31411	-0.31411	-0.30104	-0.01329	0.03622	0.06729	-0.13729	-0.21479	-0.04679
40	-0.07558	0.07559	-0.04213	0.01757	0.36837	0.05108	-0.35878	0.24114	-0.09656
41	-0.10292	0.10292	-0.05557	0.05685	0.00541	-0.06780	0.03129	0.09232	-0.03390
42	0.12093	-0.12093	0.06626	-0.04903	-0.05366	0.00557	0.27304	-0.23582	0.09276
43	-0.10697	0.10697	-0.04132	0.00155	0.10649	-0.04859	-0.23617	0.20756	-0.06038
44	-0.23396	0.23396	-0.05100	0.03846	0.03949	-0.06557	-0.13048	0.19051	0.09677
45	0.17572	-0.17572	0.03658	-0.00287	-0.09050	0.08305	0.26826	-0.27547	0.04750
46	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
47	0.59213	-0.59213	-0.08784	0.10849	-0.03821	-0.00954	0.35593	-0.31188	-0.14464
48	0.25553	-0.25553	0.08619	-0.06004	-0.07927	0.04051	0.23510	0.10414	-0.01909
49	0.19746	-0.19746	-0.08408	0.07174	0.04416	0.02997	-0.35306	-0.01562	-0.18207
50	0.06234	-0.06234	0.07304	-0.05504	-0.05667	-0.00201	0.19720	0.04829	0.10021
51	0.01821	-0.01821	0.08109	-0.08143	-0.01176	-0.10588	0.00950	0.00164	0.01459
52	0.14344	-0.14344	0.07029	-0.08107	0.01670	0.03840	0.09968	-0.10592	-0.19009
53	0.05862	-0.05862	0.06276	-0.05732	-0.02341	-0.02658	0.05291	-0.06652	-0.13153
54	0.01563	-0.01563	0.00372	-0.01349	0.02407	-0.00574	0.61358	-0.29092	0.01409
55	0.07742	-0.07742	-0.30596	0.32308	0.00457	0.03927	0.18169	-0.20883	-0.01354
56	0.03982	-0.03982	-0.12148	0.11176	0.04343	0.01089	0.07084	-0.06593	-0.16234
57	-0.46664	0.46664	0.04397	-0.07047	0.05958	-0.11954	-0.15434	0.62360	0.07973
58	0.50364	-0.50364	0.03911	-0.01040	-0.07820	0.13375	0.11514	-0.55227	-0.01631
59	0.11702	-0.11702	0.10745	-0.10129	-0.03221	0.11538	-0.01238	-0.04176	-0.08505
60	-0.12094	0.12095	-0.16667	0.15982	0.04235	-0.07164	0.05824	-0.06593	-0.05993

Table D.4 con'd

VARIABLE	VARIABLES 10	11	12	13	14	15	16	17	18
10	1.00000								
11	-0.19977	1.00000							
12	-0.87174	-0.30593	1.00000						
13	-0.08011	-0.07492	0.11530	1.00000					
14	0.13563	0.05083	-0.15721	-0.82166	1.00000				
15	-0.04345	-0.11008	0.09726	0.50863	-0.37412	1.00000			
16	-0.03332	-0.11141	0.08809	0.54566	-0.56506	0.30445	1.00000		
17	-0.03121	0.04636	0.00715	0.49379	-0.52651	0.44657	0.51041	1.00000	
18	-0.00225	-0.08084	0.04261	0.09510	-0.04814	0.19155	-0.05592	-0.38808	1.00000
19	0.07633	0.03941	-0.09387	-0.57063	0.57248	-0.60174	-0.48128	-0.63636	-0.36958
20	0.39333	-0.21050	-0.27692	0.01854	0.01658	-0.03486	-0.04695	-0.09613	0.10542
21	-0.20805	0.11634	0.14398	0.31034	-0.30289	0.26631	0.18362	0.43571	-0.20012
22	-0.26863	0.18959	0.16621	-0.21094	0.17422	-0.12217	-0.07544	-0.10148	0.06870
23	-0.00430	-0.08119	0.04478	-0.40397	0.35380	-0.25749	-0.18784	-0.26558	-0.00213
24	0.08301	-0.06777	-0.04677	-0.02803	0.01660	-0.08314	-0.06584	-0.14339	0.04259
25	0.00765	-0.11601	0.05058	-0.04632	0.04978	-0.09209	-0.04090	-0.11827	0.05686
26	0.29077	-0.04955	-0.25775	-0.03563	0.06877	-0.12967	-0.02602	-0.18170	0.05867
27	0.00349	0.01346	-0.01011	0.06642	-0.02566	0.08351	-0.01335	0.12166	-0.19516
28	0.05202	-0.12122	0.01007	0.09235	-0.06172	0.07226	0.01316	0.00440	0.08425
29	-0.00023	0.01208	-0.00582	0.19352	-0.13548	0.09812	0.06550	0.31988	-0.13769
30	0.06271	-0.13558	0.00687	0.07826	-0.08586	0.12378	0.12504	0.07303	0.02967
31	-0.18172	0.36762	-0.00726	-0.05287	0.01867	-0.08089	-0.01829	0.00514	0.02825
32	-0.33063	-0.02869	0.33560	0.00602	-0.13614	0.01179	0.01887	0.01441	0.03307
33	-0.23708	0.08497	0.18786	0.08210	-0.09283	0.18677	0.08224	0.12890	0.09467
34	0.04015	-0.00058	-0.03873	-0.08055	0.03117	-0.04426	-0.00696	-0.02612	-0.12346
35	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
36	0.01245	-0.08323	0.02952	0.17076	-0.11222	0.05437	0.03312	0.03981	0.00416
37	-0.05875	0.00921	0.05247	-0.06559	0.01642	0.00688	0.03063	-0.01351	-0.05561
38	0.11293	0.15413	-0.18680	-0.10586	0.10053	-0.08470	-0.04329	0.01642	0.10736
39	-0.16343	-0.06967	0.19363	-0.07026	0.04668	-0.09336	-0.02219	-0.17144	-0.02121
40	-0.56468	-0.12956	0.61345	-0.00360	-0.02159	0.08961	0.00557	-0.00227	0.05857
41	-0.08721	0.17256	-0.00155	0.03744	-0.04904	0.03931	-0.01968	-0.01050	0.02847
42	0.47022	-0.01449	-0.44964	-0.02115	0.04703	-0.09080	0.00842	0.00834	-0.06112
43	-0.34791	-0.04800	0.36205	0.03353	-0.03644	0.04587	-0.02283	-0.05587	0.22363
44	-0.18729	0.00749	0.17823	0.06381	-0.05393	0.07675	0.08547	0.02768	0.06527
45	0.38066	0.04460	-0.39217	-0.05523	0.05719	-0.07122	-0.04312	0.06770	-0.14650
46	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
47	0.28288	0.08956	-0.31963	-0.17093	0.14538	-0.02889	-0.05634	-0.09212	-0.08775
48	0.38092	-0.05228	-0.34397	-0.10063	0.16355	0.00846	-0.12240	-0.09028	-0.01014
49	-0.06556	-0.29462	0.21103	0.10343	-0.09328	0.13215	0.05710	-0.01804	0.04352
50	0.07922	0.12118	-0.13757	-0.11987	0.11717	-0.06751	-0.02776	0.00902	0.06688
51	0.10894	-0.03884	-0.08644	0.02365	0.00135	-0.01845	0.10748	0.03125	0.06706
52	-0.05127	0.11510	-0.00774	-0.16651	0.13218	-0.17198	-0.15020	-0.14160	-0.12127
53	0.07606	0.02746	-0.08763	-0.17129	0.22018	-0.12334	-0.12655	-0.12074	-0.03949
54	-0.11906	0.47558	-0.12212	-0.08830	0.02141	-0.08434	-0.00351	0.06277	-0.10668
55	0.03867	0.23894	-0.15705	0.01366	-0.08445	0.03593	0.01629	0.09083	-0.06858
56	-0.24836	0.25196	0.11533	-0.19051	0.12923	-0.09870	-0.00839	0.04352	-0.20298
57	-0.04254	-0.04539	0.06400	0.12720	-0.07353	0.09533	0.02684	0.14206	-0.03962
58	0.20990	-0.09188	-0.15799	-0.08858	0.05021	-0.10489	-0.05284	-0.16265	0.01042
59	-0.03080	-0.07232	0.06609	-0.31608	0.30695	-0.13333	-0.21747	-0.20513	0.00168
60	-0.24418	0.23526	0.11963	0.07020	-0.08772	0.07060	0.12581	0.11738	0.04214

Table D.4 con'd

VARIABLE	VARIABLES 19	20	21	22	23	24	25	26	27
19	1.00000								
20	0.12192	1.00000							
21	-0.24831	-0.40062	1.00000						
22	0.08066	-0.76233	-0.14351	1.00000					
23	0.27599	-0.25704	-0.06664	0.34237	1.00000				
24	0.11022	0.02454	-0.05043	0.00406	-0.00892	1.00000			
25	0.07836	0.18954	-0.11106	-0.14005	-0.04245	-0.01531	1.00000		
26	0.15635	0.31821	-0.14753	-0.24643	-0.12943	-0.04060	-0.09028	1.00000	
27	0.04016	0.04113	0.11543	-0.11585	-0.03637	-0.03306	-0.08280	-0.16914	1.00000
28	-0.05585	0.06652	-0.02432	-0.04083	-0.09215	-0.03759	-0.09378	-0.19721	-0.12888
29	-0.20598	-0.10639	0.28062	-0.05081	-0.05287	-0.02667	-0.08745	-0.12556	-0.09428
30	-0.08961	0.04209	-0.00456	-0.03564	-0.07603	-0.03739	-0.07971	-0.03602	-0.08727
31	-0.01398	-0.32185	0.03878	0.35909	-0.05038	-0.00478	-0.05656	-0.19112	-0.02133
32	-0.03538	-0.04826	0.10083	-0.00174	0.00242	-0.01212	-0.04721	-0.12202	-0.03593
33	-0.19390	-0.10808	-0.03082	0.15837	-0.09213	-0.02653	-0.09828	-0.36909	-0.16070
34	0.01969	-0.10739	0.04008	-0.02466	-0.02073	-0.03395	-0.08850	-0.19060	-0.10719
35	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
36	0.03087	0.29309	0.10133	-0.31434	-0.14555	0.06744	0.03672	0.45353	0.15580
37	-0.02367	-0.13875	-0.01714	0.07547	-0.02530	-0.05184	-0.09129	-0.34898	-0.08382
38	-0.09773	-0.17559	-0.12483	0.28691	0.14366	-0.03182	0.09666	-0.19227	-0.13115
39	0.07213	0.01438	-0.13260	-0.06666	-0.02815	0.12995	0.41329	0.06540	0.00385
40	-0.04503	-0.30479	0.12855	0.26715	0.04236	0.01098	0.01390	-0.37823	-0.06705
41	-0.03973	-0.06090	-0.01321	0.04719	-0.15010	-0.07260	0.19289	-0.16905	0.04222
42	0.05833	0.26258	-0.08603	-0.22622	0.06427	0.03807	-0.13279	0.38526	0.02243
43	-0.09834	-0.32264	-0.06360	0.44018	-0.06494	-0.01977	-0.07486	-0.27638	-0.06193
44	-0.03241	-0.10949	0.07636	0.13202	0.01164	0.00306	-0.07728	-0.32063	-0.05536
45	0.13019	0.37258	0.02181	-0.34969	0.04230	0.01181	0.10025	0.39152	0.07740
46	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
47	0.16578	0.15897	-0.13171	-0.09691	0.02718	0.07967	0.17299	0.15155	0.12215
48	0.09524	0.21881	-0.10084	-0.19077	-0.03448	0.09356	0.04004	0.08189	0.06095
49	-0.00544	0.26434	0.04394	-0.31615	-0.13068	0.06709	0.03887	0.18181	0.06813
50	-0.05736	-0.12959	-0.13462	0.24041	0.05438	-0.03544	0.09868	-0.19698	-0.16710
51	-0.08292	0.04000	-0.10528	0.02486	-0.00066	-0.02090	-0.00831	-0.06210	-0.06203
52	0.22437	-0.10011	0.15090	-0.00134	0.26242	-0.02123	-0.05932	-0.05071	0.25068
53	0.15635	-0.05769	-0.09064	0.12565	0.38092	-0.00408	0.00953	0.02329	-0.04494
54	0.03913	-0.25442	0.10247	0.24308	0.02101	-0.06616	-0.02855	-0.09848	0.08457
55	-0.03098	0.02091	-0.06666	0.02376	-0.09563	-0.02516	-0.00623	0.01515	0.04864
56	0.11656	-0.25903	-0.01070	0.29656	0.14726	-0.03927	0.04562	-0.15213	-0.03226
57	-0.09495	-0.01565	0.14774	-0.05730	-0.06046	-0.06880	-0.23833	-0.12109	-0.11373
58	0.12646	0.27693	-0.24674	-0.19319	-0.09978	0.07352	0.25949	0.33707	0.15822
59	0.21113	-0.23930	-0.07112	0.32420	0.83066	-0.00520	-0.03563	-0.10046	-0.03267
60	-0.13445	-0.30426	0.18439	0.25167	-0.08409	-0.00792	-0.08868	-0.29401	-0.05946

Table D.4 con'd

VARIABLE	VARIABLES 28	29	30	31	32	33	34	35	36
28	1.00000								
29	-0.06485	1.00000							
30	0.18242	-0.02590	1.00000						
31	-0.12312	-0.04875	-0.05422	1.00000					
32	-0.06141	-0.03630	0.00343	0.01609	1.00000				
33	-0.18763	-0.15433	-0.15319	-0.09425	0.00848	1.00000			
34	-0.15556	-0.05594	-0.05297	-0.05546	-0.02880	-0.12233	1.00000		
35	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
36	0.15897	0.11628	0.04599	0.00045	-0.00933	-0.32231	-0.65743	0.0	1.00000
37	-0.25324	-0.13205	-0.13477	-0.09939	0.01906	0.33832	0.77436	0.0	-0.81798
38	0.16416	0.02423	0.16051	0.18781	-0.00883	0.00555	-0.14463	0.0	-0.31965
39	-0.04744	-0.04430	-0.04894	-0.07941	0.05589	-0.16358	-0.04808	0.0	0.11936
40	0.02926	-0.00965	0.00618	0.12795	0.40068	0.24317	-0.00583	0.0	-0.10326
41	-0.02132	-0.11897	-0.04126	-0.02830	-0.03254	0.32941	-0.11898	0.0	-0.14567
42	-0.00794	0.08270	0.02168	-0.07599	-0.27364	-0.38797	0.07990	0.0	0.16842
43	0.08156	-0.06667	-0.01215	0.07426	0.29383	0.26383	0.01970	0.0	-0.18798
44	0.12597	-0.00605	0.04219	0.04272	0.25554	0.27184	-0.06622	0.0	-0.10258
45	-0.11494	0.05282	-0.00950	-0.06407	-0.34408	-0.32710	-0.06105	0.0	0.24753
46	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
47	-0.08751	-0.14697	0.03184	0.07335	-0.09600	-0.27066	0.04941	0.0	0.03040
48	-0.02488	-0.00948	-0.02814	-0.10363	-0.17025	-0.04702	0.02906	0.0	-0.06400
49	-0.07875	-0.06310	-0.08010	-0.14643	0.03091	-0.09952	0.10443	0.0	0.31412
50	0.12962	0.00511	0.16946	0.11248	-0.00461	0.12148	-0.10597	0.0	-0.34742
51	0.18396	0.01052	0.06906	-0.01329	-0.02397	0.01333	-0.06999	0.0	0.01220
52	-0.09870	-0.05537	-0.06864	-0.03789	-0.02027	-0.10380	0.08815	0.0	-0.06556
53	-0.06605	-0.02524	-0.04904	-0.02678	-0.02046	-0.04214	-0.04205	0.0	-0.03697
54	-0.09051	0.02219	0.02059	0.30433	0.24710	-0.09175	-0.02241	0.0	-0.06698
55	0.00248	-0.08430	-0.03612	0.05620	0.04537	0.00711	-0.02162	0.0	-0.00480
56	-0.04047	-0.02006	-0.03610	0.30037	0.06210	0.04917	-0.10377	0.0	-0.04955
57	0.03347	0.27825	-0.08743	-0.10551	-0.05267	0.20219	0.12214	0.0	-0.13187
58	0.02168	-0.23995	0.15753	-0.13466	-0.06677	-0.28900	-0.21615	0.0	0.26568
59	-0.08369	-0.03000	-0.04958	-0.05596	-0.02899	-0.06247	-0.03394	0.0	-0.10334
60	-0.04973	-0.03526	-0.09072	0.38291	0.19088	0.16465	0.16119	0.0	-0.16834

Table D.4 con'd

VARIABLE	VARIABLES 37	38	39	40	41	42	43	44	45
37	1.00000								
38	-0.23774	1.00000							
39	-0.10851	-0.04223	1.00000						
40	0.15373	-0.08228	0.12025	1.00000					
41	0.14477	0.03692	0.10684	0.06058	1.00000				
42	-0.20493	0.03699	-0.15623	-0.77306	-0.68000	1.00000			
43	0.20993	0.00406	-0.09279	0.54134	0.21406	-0.53368	1.00000		
44	0.08934	-0.04269	-0.05170	0.37175	0.32586	-0.48015	0.21593	1.00000	
45	-0.26142	0.02728	-0.00143	-0.58506	-0.35715	0.65673	-0.78559	-0.68944	1.00000
46	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
47	-0.08162	0.11506	0.03953	-0.42708	-0.19323	0.43652	-0.32307	-0.38529	0.44524
48	-0.08227	0.29502	-0.03805	-0.39078	-0.22225	0.42829	-0.29669	-0.24167	0.33952
49	0.16970	-0.76047	0.06447	0.06589	-0.13219	0.03561	0.05257	-0.09072	0.02635
50	-0.15951	0.91479	-0.07353	-0.03415	0.10212	-0.03982	0.06007	0.02188	-0.04060
51	-0.07245	0.10865	0.00871	0.01661	0.00556	-0.01574	0.07406	0.08011	-0.09437
52	0.05656	0.02573	-0.07018	-0.01680	0.13331	-0.07237	-0.06347	-0.13970	0.11032
53	-0.06103	0.17323	-0.01404	-0.08019	-0.08361	0.11204	-0.02213	-0.10267	0.07647
54	-0.05964	0.21363	-0.02213	0.08009	0.12960	-0.14119	0.10059	0.10820	-0.11918
55	0.00807	0.02095	0.11543	-0.02998	0.12289	-0.05607	-0.03112	0.09670	-0.03113
56	-0.06394	0.10521	-0.02307	0.15459	0.09464	-0.17370	0.06950	0.13809	-0.13215
57	0.19239	-0.07833	-0.16495	0.03431	-0.07436	0.02203	0.10519	0.03313	-0.07805
58	-0.30482	0.06962	0.24177	-0.22992	0.08581	0.11437	-0.27088	-0.17121	0.26086
59	-0.03927	0.05674	-0.02970	0.05196	-0.12621	0.04205	-0.06088	0.02785	0.02957
60	0.19414	-0.01215	-0.11185	0.27852	0.02901	-0.22301	0.28008	0.20043	-0.29275

VARIABLE	VARIABLES 46	47	48	49	50	51	52	53	54
46	0.0								
47	0.0	1.00000							
48	0.0	0.51151	1.00000						
49	0.0	0.12591	-0.10692	1.00000					
50	0.0	0.08136	0.33490	-0.74683	1.00000				
51	0.0	0.01214	0.09081	-0.00841	0.06091	1.00000			
52	0.0	0.12653	-0.00346	-0.03312	-0.03685	-0.06669	1.00000		
53	0.0	0.12066	0.09635	-0.01545	0.00233	-0.02753	0.35357	1.00000	
54	0.0	0.18942	-0.07098	-0.36836	0.17112	-0.02535	0.16616	0.01044	1.00000
55	0.0	0.15369	-0.02042	-0.03313	0.01886	-0.09031	-0.03405	-0.00924	0.11445
56	0.0	0.23935	-0.17333	-0.14083	0.10406	-0.04348	0.01894	0.05610	0.28464
57	0.0	-0.37176	0.02655	-0.02398	-0.05493	-0.01115	-0.08640	-0.08214	-0.26054
58	0.0	0.36999	0.09117	0.01911	0.07763	0.03832	0.04008	-0.02127	0.10119
59	0.0	0.05188	-0.00441	-0.05328	-0.01819	-0.02453	0.33242	0.46094	0.02682
60	0.0	-0.03264	-0.17624	0.02730	-0.02962	-0.03216	-0.06324	-0.02658	0.22131

VARIABLE	VARIABLES 55	56	57	58	59	60
55	1.00000					
56	0.14622	1.00000				
57	-0.19236	-0.23946	1.00000			
58	0.16844	0.04083	-0.75704	1.00000		
59	-0.08735	0.17352	-0.09305	-0.11847	1.00000	
60	0.06283	0.22445	-0.29265	-0.35368	-0.07374	1.00000

Appendix E

Conceptual Approach:

Discriminant Analysis With

An Initial List of 59 Variables

Notation

a. Tables

1. INDPRI = Independent Primary Segment

2. PRISEC = Subordinate Primary Segment

3. SECSEC = Secondary Segment

b. Figures

1. I = Independent Primary Segment

2. P = Subordinate Primary Segment

3. S = Secondary Segment

4. * = Segment Means

5. £ or ∯ = Point of Overlap

Table E.1.a

Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE *	NUMBER OF VARIABLES INCLUDED
1	47		223.6713	1
2	8		163.2101	2
3	56		33.7693	3
4	55		6.1411	4
5	26		5.9133	5
6	37		3.6906	6
7	7		5.3105	7
8	1		10.6595	8
9	60		5.2412	9
10	40		5.0905	10
11	16		3.5381	11
12	54		3.2100	12
13	3		2.9089	13
14	39		2.9558	14
15	28		2.3330	15
16	45		2.7826	16
17	34		2.5183	17
18	32		2.4790	18
19	9		1.7957	19
20	29		1.7316	20
21	49		1.5271	21
22	10		1.6886	22
23	5		1.3119	23
24	23		1.1308	24
25	36		1.5256	25
26	15		1.7790	26
27	53		1.5307	27
28	38		1.2920	28
29	14		1.0464	29
30	33		0.8866	30
31	59		0.7060	31
32	13		0.7212	32
33	6		0.9061	33
34	52		0.4899	34
35	11		0.4774	35
36	24		0.4160	36
37	57		0.3533	37
38	21		0.2571	38
39	30		0.2471	39
40	19		0.2400	40
41	51		0.2419	41
42	50		0.1518	42
43	25		0.1939	43
44	42		0.1355	44
45	22		0.0816	45
46	43		0.1503	46
47	20		0.0131	47

* F-ratio at .01 level with 47 and 246 degrees of freedom ≥ 1.61

Table E.1.b

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	————	9.365	28.692
Subordinate	9.365	————	15.466
Secondary	28.692	15.466	————
F - Statistic**	13.144		
* F-ratio at .01 level with 47 and 246 degrees of freedom 1.61			
** F-ratio at .01 level with 94 and 492 degrees of freedom 1.42			

Table E.1.c

Discrminant Functions

	FUNCTION		
VARIABLE	INDPRI	PRISUB	SECSEC
1	2.03383	1.96759	1.89637
3	15.71974	15.84426	15.64497
5	15.76810	15.51725	15.25377
6	-2.77956	-2.85063	-3.10911
7	3.75463	3.27377	3.21158
8	33.37399	30.43063	29.06471
9	0.01864	0.03427	0.03213
10	0.61431	0.57821	0.53867
11	2.88071	2.94614	2.91453
13	2.23623	2.21124	2.14390
14	2.36213	2.41289	2.37128
15	0.01235	0.04580	0.10001
16	2.27693	2.38328	2.39049
19	0.53487	0.51412	0.51027
20	5.38546	5.38499	5.39575
21	5.88805	5.87857	5.91782
22	6.19046	6.17648	6.20065
23	-1.19836	-1.24034	-1.20449
24	1.43503	1.49130	1.50653
25	0.04276	0.05642	0.06962
26	-0.34083	-0.34324	-0.29756
28	-0.55183	-0.50809	-0.48403
29	-0.24510	-0.21983	-0.22453
30	0.05392	0.08571	0.08435
32	-0.03237	-0.00664	-0.07707
33	-0.37049	-0.37295	-0.35010
34	0.48473	0.43001	0.46554
36	5.62714	5.72584	5.91680
37	5.45920	5.68474	5.86107
38	5.99899	6.05428	6.21222
39	3.10154	2.90383	3.12783
40	0.57407	0.59281	0.68404
42	1.33850	1.34523	1.36663
43	-0.17683	-0.16255	-0.18254
45	-0.24224	-0.19674	-0.24861
47	-2.40562	-2.38222	-2.47549
49	-0.15791	-0.29444	-0.40299
50	-1.47665	-1.52415	-1.55797
51	-4.04095	-4.28853	-4.31880
52	-2.63319	-2.81347	-2.65656
53	-19.71773	-19.09383	-19.44379
54	-1.96356	-1.84597	-1.86461
55	6.87394	6.88297	7.06485
56	3.37869	3.36236	3.59316
57	-36.55849	-37.30679	-37.99274
59	205.55684	212.68037	216.55515
60	-5.83913	-9.33389	-10.36620
CONSTANT	-1852.49487	-1799.31982	-1757.42334

Table E.1.d

Classification Matrix

Segment Before the Analysis	Predicted Segment			Total
	Independent	Subordinate	Secondary	
Independent	54	3	0	57
Subordinate	8	110	5	123
Secondary	0	3	112	115
Previously Unclassified	28	57	48	133
Total Classified	90	173	165	428

Table E.1.e

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	57	19	33.3	10	17.5
Subordinate	123	69	56.1	13	10.6
Secondary	115	60	52.2	23	20.0
Total	295	148	50.2	46	15.6

* Chi^2 at .10 level with 47 degrees of freedom $\leq$ 34.92

** Chi^2 at .99 level with 47 degrees of freedom $\geq$ 73.51

Figure E.1

Two-Dimensional Representation

of the Labour Market Segments

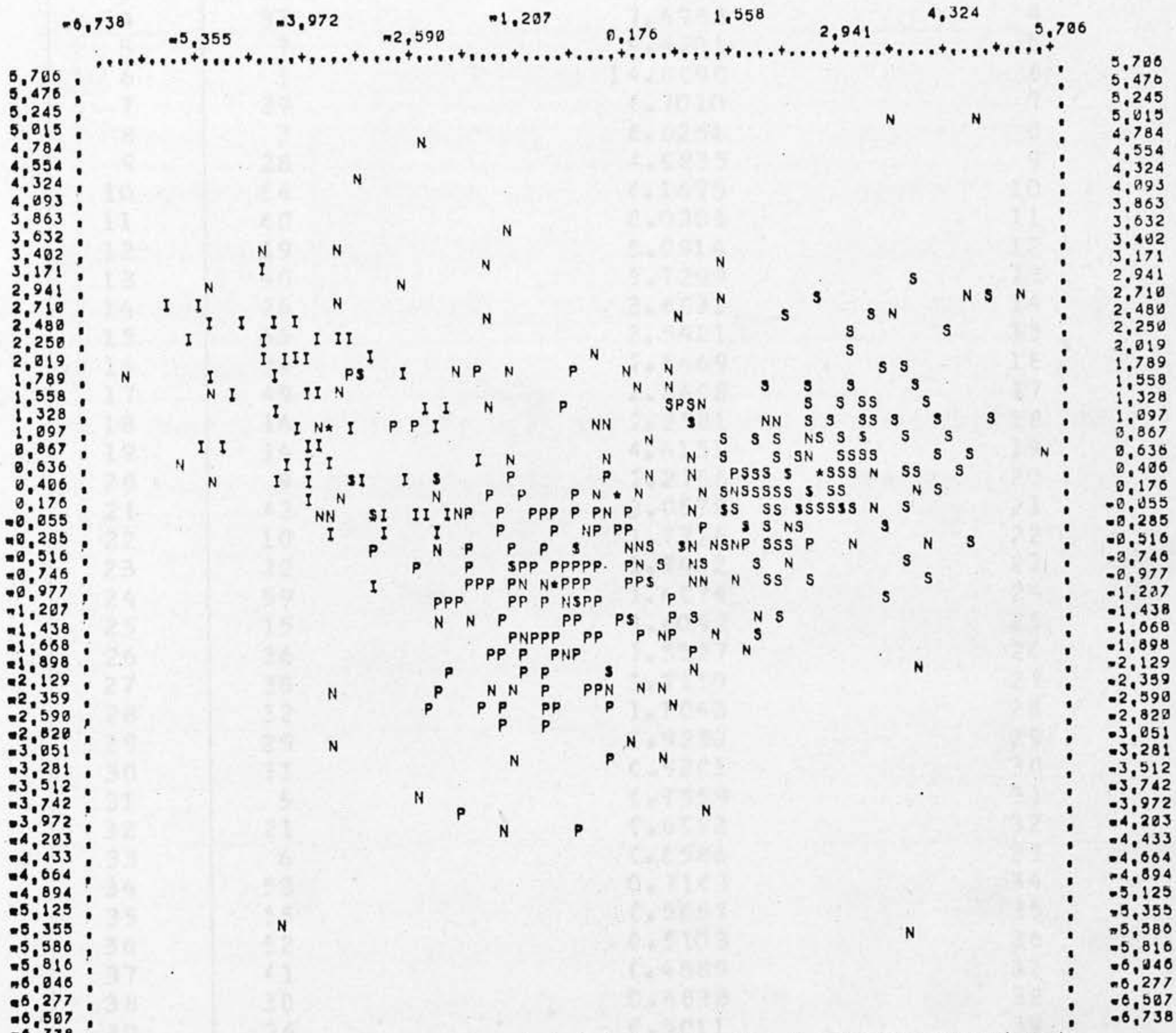


Table E.2.a

Summary Table

STEP NUMBER	VARIABLE ENTERED	REMOVED	F VALUE TO ENTER OR REMOVE*	NUMBER OF VARIABLES INCLUDED
1	47		304.5088	1
2	8		194.9636	2
3	56		27.1595	3
4	37		7.6983	4
5	7		9.4201	5
6	1		14.8090	6
7	39		8.7010	7
8	3		6.8251	8
9	28		4.9835	9
10	54		6.1675	10
11	60		8.0201	11
12	19		5.0914	12
13	40		3.7299	13
14	26		3.6931	14
15	45		3.5421	15
16	34		2.5469	16
17	49		2.8408	17
18	16		2.2581	18
19	14		4.4151	19
20	9		2.2796	20
21	43		2.0578	21
22	10		1.7716	22
23	22		1.7442	23
24	59		1.6074	24
25	15		1.6092	25
26	36		1.5527	26
27	38		1.2150	27
28	32		1.1043	28
29	29		0.9230	29
30	31		0.9201	30
31	5		0.7559	31
32	21		0.6982	32
33	6		0.6586	33
34	53		0.7163	34
35	55		0.5657	35
36	52		0.5103	36
37	41		0.4889	37
38	30		0.4838	38
39	24		0.5011	39
40	51		0.3840	40
41	11		0.3120	41
42	58		0.3063	42
43	20		0.3206	43
44	50		0.2769	44
45	25		0.2490	45
46	27		0.1583	46
47	13		0.0813	47
48	23		0.0274	48
49	18		0.0105	49

* F-ratio at .01 level with 49 and 225 degrees of freedom $\geq$ 1.60

Table E.2.b

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	----------	13.324	40.418
Subordinate	13.324	----------	20.694
Secondary	40.418	20.694	---------
F - Statistic**	17.641		

* F-ratio at .01 level with 49 and 225 degrees of freedom 1.60

** F-ratio at .01 level with 98 and 450 degrees of freedom 1.43

Table E.2.c

Discrminant Functions

VARIABLE	FUNCTION INDPRI	PRISUB	SECSEC
1	2.65333	2.51899	2.43252
3	17.87596	17.92435	17.62459
5	18.69481	18.49252	18.58568
6	-1.84098	-1.83588	-2.13953
7	5.59577	4.83915	4.67699
8	47.58261	42.90045	41.16377
9	0.08174	0.09273	0.07565
10	0.63155	0.56737	0.56636
11	3.03614	3.10734	3.12423
13	2.50047	2.48033	2.45738
14	2.59671	2.70198	2.80678
15	-0.10294	-0.04915	0.00209
16	2.73746	2.89847	3.00684
18	0.35980	0.36541	0.36395
19	1.06699	1.00042	0.97878
20	5.04515	5.05722	5.13202
21	5.55645	5.57208	5.69925
22	5.92098	5.88088	6.02677
23	-0.70396	-0.70833	-0.72142
24	2.16665	2.24056	2.24004
25	0.42277	0.41581	0.39513
26	0.28489	0.28164	0.30875
27	0.75828	0.77358	0.77105
28	0.02214	0.09274	0.09347
29	0.66308	0.64657	0.60838
30	0.68249	0.71698	0.68615
31	1.29394	1.28930	1.25325
32	0.31800	0.44186	0.54222
34	1.20913	1.17441	1.21585
36	5.30202	5.37643	5.64636
37	4.89832	5.13124	5.37564
38	5.11508	5.12985	5.40548
39	3.05061	2.99736	3.57915
40	-0.27156	-0.24161	-0.14050
41	0.38075	0.43415	0.44230
43	0.00548	0.05498	-0.05962
45	0.39304	0.49637	0.43519
47	-2.83752	-2.77894	-3.00355
49	-0.61384	-0.76628	-0.89137
50	-1.50258	-1.50049	-1.57734
51	-2.36365	-2.75685	-3.02118
52	-7.21731	-7.58881	-7.46396
53	-11.77161	-10.74728	-11.43196
54	-3.24484	-3.03436	-3.04212
55	6.99088	6.98886	7.12819
56	2.77204	2.72025	2.90379
58	50.19476	50.59924	52.09700
59	263.06030	270.95703	280.82300
60	52.84706	48.52274	49.44160
CONSTANT	-2124.16943	-2026.79761	-1978.77710

Table E.2.d

Classification Matrix

Segment Before the Analysis	Predicted Segment			
	Independent	Subordinate	Secondary	Total
Independent	54	0	0	54
Subordinate	1	109	0	110
Secondary	0	0	112	112
Previously Unclassified	36	64	52	152
Total Classified	91	173	164	428

Table E.2.e

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	54	20	37.0	8	14.8
Subordinate	110	60	54.5	13	11.8
Secondary	112	59	52.7	21	18.8
Total	276	139	50.4	42	15.2

* Chi^2 at .10 level with 49 degrees of freedom $\leq$ 36.70

** Chi^2 at .99 level with 49 degrees of freedom $\geq$ 75.64

Figure E.2

Two-Dimensional Representation of the Labour Market Segments

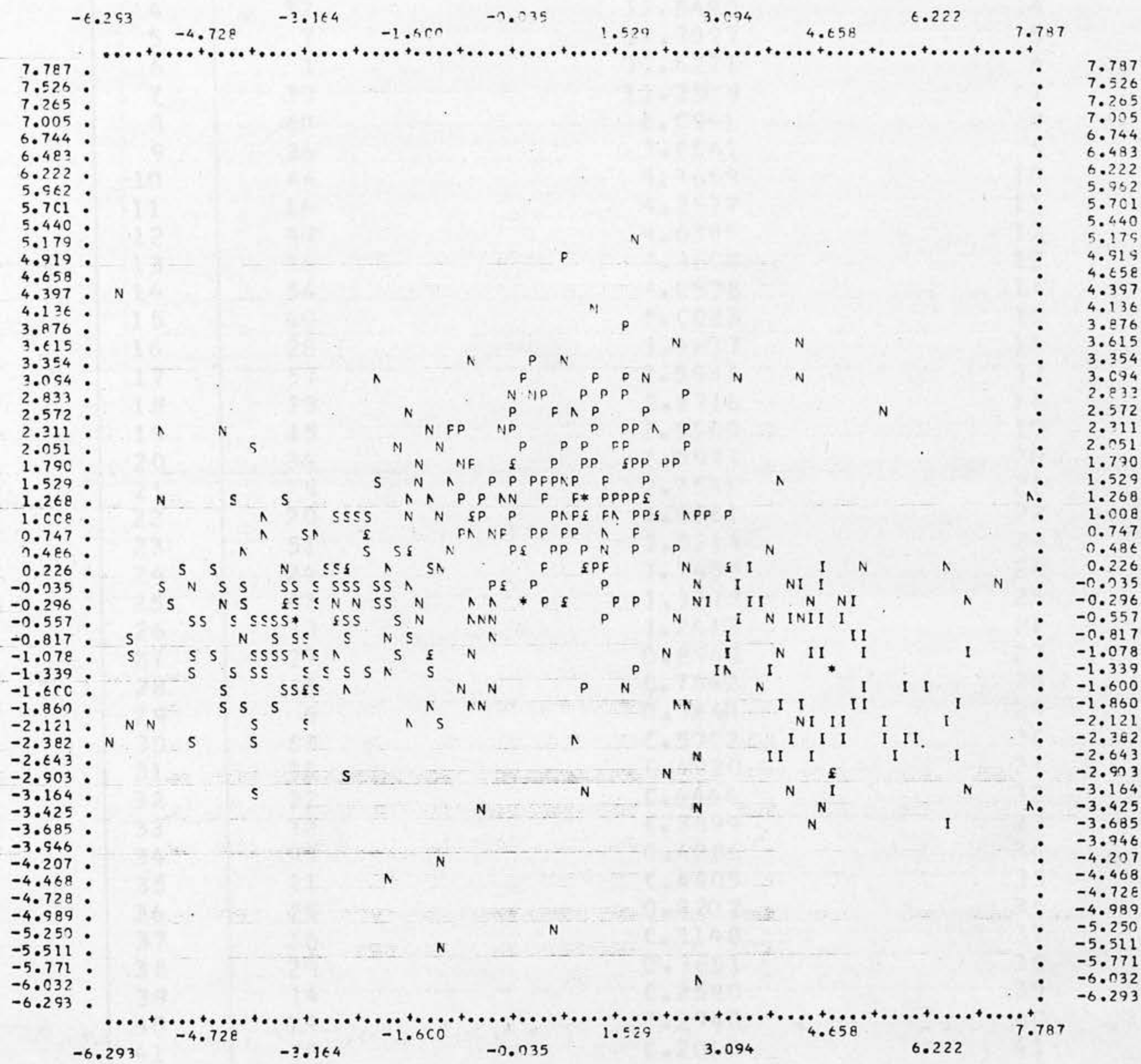


Table E.3.a

Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE*	NUMBER OF VARIABLES INCLUDED
1	47		349.7844	1
2	8		182.3623	2
3	3		11.7388	3
4	37		12.5490	4
5	7		14.7077	5
6	1		13.6271	6
7	39		12.3509	7
8	60		6.0941	8
9	26		7.8541	9
10	44		4.3699	10
11	16		4.2577	11
12	49		4.6385	12
13	36		3.3608	13
14	54		4.0578	14
15	40		5.0023	15
16	28		3.0607	16
17	57		2.5987	17
18	23		2.8716	18
19	15		3.5105	19
20	34		3.0977	20
21	9		2.3531	21
22	20		1.9561	22
23	51		2.0214	23
24	24		1.7458	24
25	43		1.3378	25
26	13		1.2413	26
27	27		0.8948	27
28	5		0.7842	28
29	6		0.7640	29
30	53		0.5702	30
31	33		0.4920	31
32	22		0.4444	32
33	32		0.3899	33
34	55		0.4286	34
35	11		0.4405	35
36	25		0.3202	36
37	56		0.3148	37
38	29		0.3651	38
39	14		0.2580	39
40	19		0.2748	40
41	38		0.2010	41
42	18		0.2198	42
43	12		0.1197	43
44	50		0.0964	44
45	31		0.0709	45
46	41		0.0660	46
47	52		0.0115	47

* F-ratio at .01 level with 47 and 228 degrees of freedom ≥ 1.61

Table E.3.b

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	----------	14.118	39.460
Subordinate	14.118	----------	21.684
Secondary	39.460	21.684	----------
F - Statistic**	18.384		

* F-ratio at .01 level with 47 and 228 degrees of freedom 1.61

** F-ratio at .01 level with 94 and 456 degrees of freedom 1.42

Table E.3.c

Discrminant Functions

VARIABLE	FUNCTION INDPRI	PRISUB	SECSEC
1	1.49133	1.31561	1.17164
3	27.91216	27.83977	27.34610
5	25.66507	25.34674	25.40659
6	-0.62054	-0.77766	-0.65804
7	8.12109	7.28905	7.27114
8	43.63403	39.27533	38.04556
9	-0.25362	-0.23665	-0.25673
11	1.00614	1.03303	0.95586
12	1.90099	1.92292	1.94372
13	9.43843	9.43048	9.34935
14	9.08090	9.13698	9.10900
15	-1.42647	-1.39363	-1.30767
16	7.73712	7.89053	7.95426
18	3.84558	3.84911	3.88398
19	4.37537	4.34160	4.36554
20	0.78953	0.83382	0.85590
22	7.89484	7.79305	7.75958
23	-2.70542	-2.53342	-2.19510
24	-4.15932	-3.91262	-5.41891
25	2.24820	2.20543	2.22718
26	1.31638	1.27937	1.33334
27	2.35647	2.34712	2.39906
28	2.39085	2.42127	2.45471
29	3.77092	3.71827	3.72027
31	1.22953	1.20923	1.21313
32	9.97585	9.94011	10.14056
33	3.29493	3.28217	3.31327
34	5.76650	5.74190	5.87360
36	211.77919	211.86293	212.41321
37	207.89163	208.12512	208.59158
38	217.71188	217.84390	218.21666
39	-3.31389	-3.14748	-2.19076
40	3.84667	3.78751	3.88420
41	1.13927	1.16056	1.16871
43	-5.08819	-4.99762	-4.96142
44	-0.24673	-0.39126	-0.34642
47	1.11682	1.12068	0.92243
49	-2.05586	-2.09995	-2.35661
50	-10.52266	-10.51833	-10.46656
51	-7.37329	-8.55280	-9.12272
52	-5.78659	-5.89771	-5.96307
53	110.50844	109.01379	108.09744
54	-7.06610	-6.77004	-6.94513
55	29.00868	29.07672	29.35310
56	4.87454	4.76960	4.74239
57	-50.59113	-52.37021	-55.85817
60	-42.29027	-47.55507	-48.55206
CONSTANT	-13257.39062	-13145.82031	-13100.89453

Table E.3.d

Classification Matrix

Segment Before the Analysis	Predicted Segment Independent	Subordinate	Secondary	Total
Independent	52	0	0	52
Subordinate	1	119	0	120
Secondary	0	0	105	105
Previously Unclassified	38	51	62	151
Total Classified	91	170	167	428

Table E.3.e

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	52	17	32.7	14	26.9
Subordinate	120	55	45.8	17	14.2
Secondary	105	55	52.4	18	17.1
Total	277	127	45.8	49	17.7

* Chi^2 at .10 level with 47 degrees of freedom $\leq$ 34.92

** Chi^2 at .99 level with 47 degrees of freedom $\geq$ 73.51

Figure E.3

Two-Dimensional Representation

of the Labour Market Segments

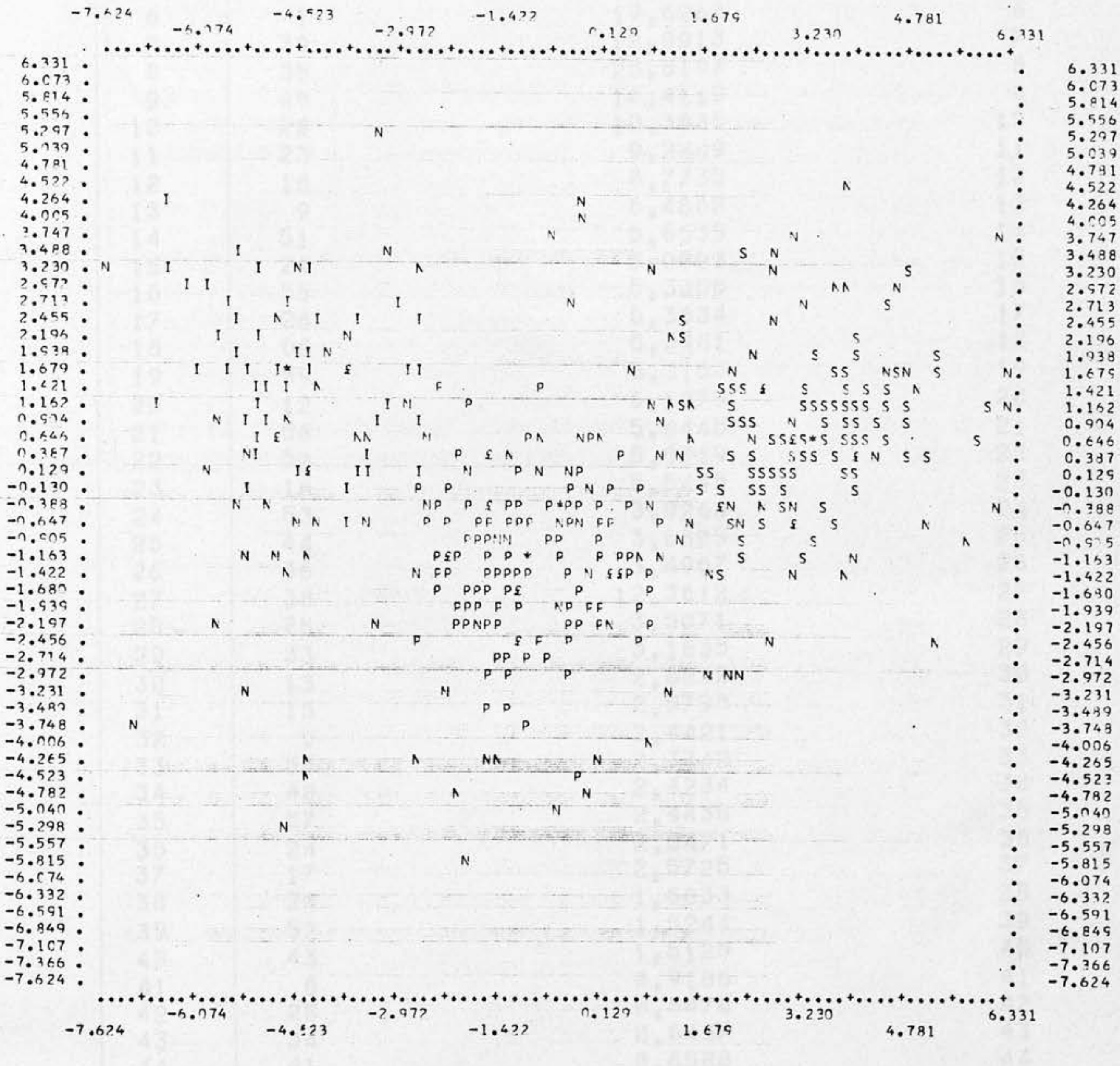


Table E.4.a
Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE*	NUMBER OF VARIABLES INCLUDED
1	47		215.4340	1
2	8		199.2477	2
3	3		21.0723	3
4	37		14.1623	4
5	7		15.9804	5
6	1		19.6268	6
7	39		12.9913	7
8	35		25.5197	8
9	49		16.4119	9
10	22		10.3835	10
11	23		9.2449	11
12	16		8.7735	12
13	9		6.4862	13
14	51		5.6535	14
15	29		5.9023	15
16	55		5.3356	16
17	26		5.3634	17
18	60		5.3481	18
19	59		5.3153	19
20	12		6.1279	20
21	56		5.4445	21
22	50		5.5019	22
23	18		5.8445	23
24	53		3.9268	24
25	44		3.6805	25
26	36		3.4967	26
27	38		12.3612	27
28	25		3.5071	28
29	33		3.1835	29
30	13		2.8032	30
31	15		2.5798	31
32	5		2.4421	32
33	54		2.3342	33
34	40		2.4234	34
35	57		2.4430	35
36	24		2.0471	36
37	17		2.5725	37
38	20		1.6633	38
39	52		1.5244	39
40	43		1.5125	40
41	6		0.9186	41
42	28		0.8478	42
43	34		0.6437	43
44	41		0.6588	44
45	10		0.5748	45
46	30		0.5146	46
47	32		0.4592	47
48	19		0.2353	48
49	14		0.1025	49
50	31		0.0437	50
51	42		0.0341	51

* F-ratio at .01 level with 51 and 375 degrees of freedom ≥ 1.58

Table E.4.b

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	----------	16.894	42.002
Subordinate	16.894	----------	19.851
Secondary	42.002	19.851	----------
F - Statistic**	20.583		

* F-ratio at .01 level with 51 and 375 degrees of freedom 1.58

** F-ratio at .01 level with 102 and 750 degrees of freedom 1.40

Table E.4.c

Discriminant Functions

VARIABLE	FUNCTION INDPRI	PRISUB	SECSEC
1	0.92292	0.82326	0.73425
3	2.80000	2.78711	2.54498
5	2.38530	2.02240	1.83175
6	-0.88723	-0.89087	-0.71375
7	4.72948	4.35068	4.39136
8	25.35371	22.24670	21.76265
9	0.10498	0.09875	0.05980
10	0.56444	0.59352	0.60350
12	0.78055	0.85483	0.88703
13	1.00199	0.95723	0.90644
14	0.71957	0.72688	0.71374
15	0.21264	0.22211	0.25994
16	0.75115	0.82367	0.89540
17	-1.36734	-1.28858	-1.27862
18	-0.97047	-0.85687	-0.80701
19	-0.75134	-0.72542	-0.71136
20	0.62662	0.58734	0.60578
22	1.43876	1.33159	1.31053
23	-0.39631	-0.28428	-0.11009
24	-0.05989	0.02468	0.02401
25	0.71028	0.67292	0.65672
26	0.29779	0.26577	0.27762
28	0.17925	0.19584	0.19468
29	0.46230	0.37861	0.37484
30	-0.08938	-0.06045	-0.07669
31	0.27833	0.27152	0.26906
32	0.35726	0.38633	0.40049
33	-0.00346	-0.02148	-0.03346
34	0.25369	0.25811	0.28535
35	5.17311	4.91134	4.31116
36	6.37322	6.44980	6.72017
37	6.02754	6.19208	6.43494
38	5.04291	5.18144	5.40702
39	-1.01834	-0.94319	-0.39377
40	4.51189	4.50681	4.56417
41	4.27184	4.28086	4.31116
42	4.69827	4.68396	4.69020
43	-0.42347	-0.43948	-0.41146
44	0.37462	0.30915	0.31878
47	-0.58399	-0.54530	-0.68132
49	-0.59270	-0.75175	-0.88391
50	0.00768	-0.15689	-0.18205
51	0.51157	-0.06789	-0.22874
52	-2.73600	-2.82029	-2.73062
53	3.68804	3.25587	2.93723
54	0.39685	0.42542	0.33747
55	0.90375	0.95589	1.13885
56	0.46408	0.37886	0.40307
57	-11.57697	-12.43730	-14.32932
59	185.33083	175.58844	164.79524
60	29.82205	25.76003	25.02950
CONSTANT	-1060.44312	-984.10278	-963.45435

Table E.4.d

Classification Matrix

Segment Before the Analysis	Predicted Segment Independent	Subordinate	Secondary	Total
Independent	87	4	0	91
Subordinate	1	166	3	170
Secondary	0	1	166	167
Previously Unclassified	0	0	0	0
Total Classified	88	171	169	428

Table E.4.e

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	91	45	49.5	17	18.7
Subordinate	170	114	67.1	15	8.8
Secondary	167	99	59.3	33	19.8
Total	428	258	60.3	65	15.2

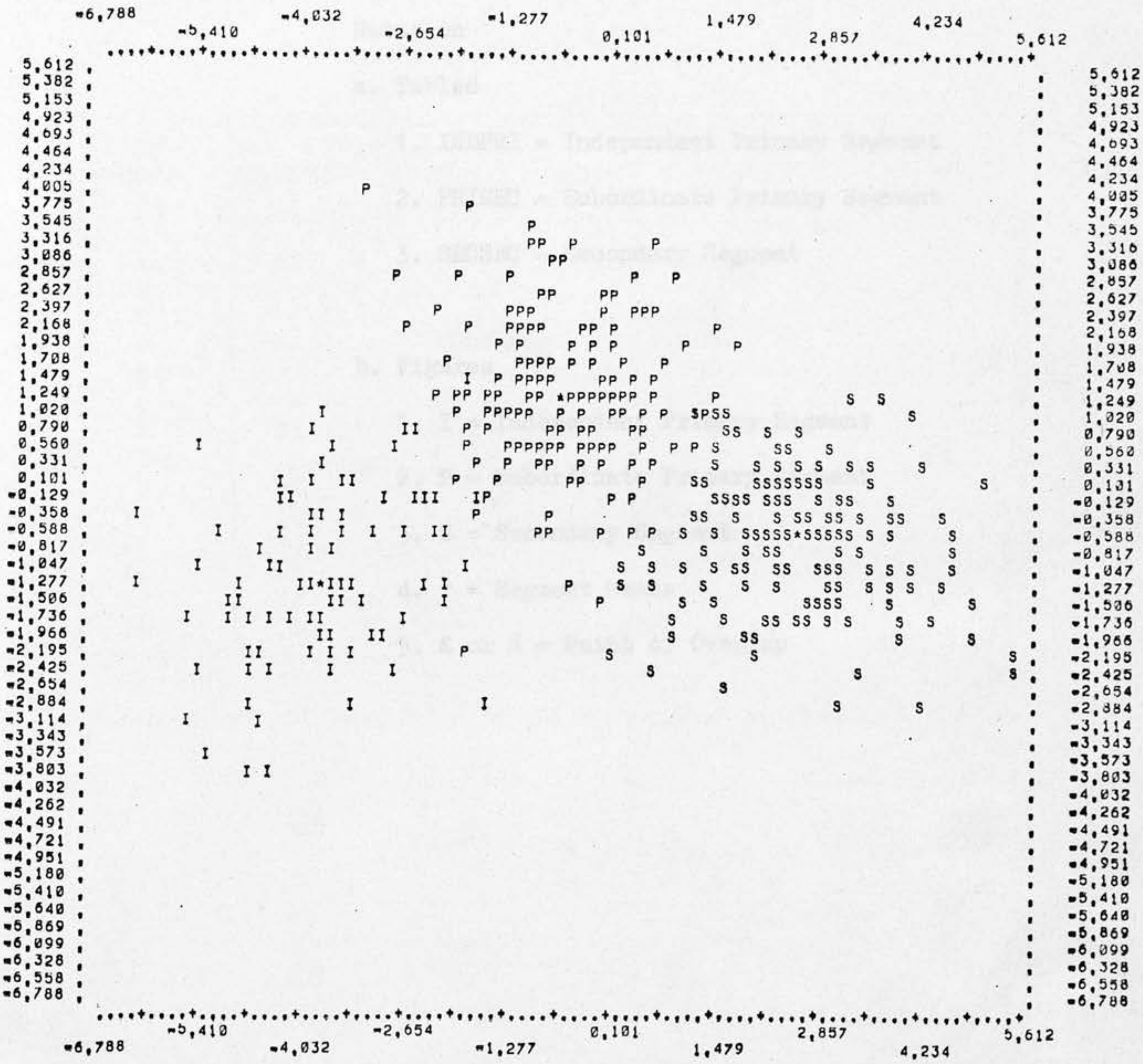
* Chi^2 at .10 level with 51 degrees of freedom $\leq$ 38.40

** Chi^2 at .99 level with 51 degrees of freedom $\geq$ 77.21

Figure E.4

Two-Dimensional Representation

of the Labour Market Segments



Appendix F

Conceptual Approach:

Discriminant Analysis With

An Initial List of 16 Variables

Notation

a. Tables

1. INDPRI = Independent Primary Segment
2. PRISEC = Subordinate Primary Segment
3. SECSEC = Secondary Segment

b. Figures

1. I = Independent Primary Segment
2. P = Subordinate Primary Segment
3. S = Secondary Segment
4. * = Segment Means
5. £ or ♯ = Point of Overlap

Table F.1.a

Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE*	NUMBER OF VARIABLES INCLUDED
1	48		587.0740	1
2	8		29.0903	2
3	1		24.3059	3
4	56		13.6044	4
5	7		6.6995	5
6	4		4.4853	6
7	31		2.1132	7
8	3		1.7996	8
9	58		1.2795	9
10	33		1.1821	10
11	23		0.8831	11
12	59		0.9474	12
13	57		0.1580	13
14	6		0.0622	14

* F-ratio at .01 level with 14 and 279 degrees of freedom $\geq$ 1.67

Table F.1.b

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	----------	38.542	107.183
Subordinate	38.542	----------	38.061
Secondary	107.183	38.061	----------
F - Statistic**	39.231		

* F-ratio at .01 level with 14 and 279 degrees of freedom 1.67

** F-ratio at .01 level with 28 and 558 degrees of freedom 1.72

Table F.1.c

Discriminant Functions

VARIABLE	FUNCTION INDPRI	PRISEC	SECSEC
1	Ø.53647	Ø.51236	Ø.42547
3	54.66479	55.Ø5562	55.1561Ø
4	53.59286	53.98581	54.27763
6	1.93615	1.899Ø8	1.85224
7	2.32645	2.3Ø678	2.15514
8	29.12515	27.97174	26.67Ø78
23	-Ø.687Ø5	-Ø.73568	-Ø.725Ø8
31	Ø.26954	Ø.24671	Ø.23Ø81
33	-Ø.3Ø876	-Ø.31387	-Ø.3Ø196
48	-Ø.6Ø94Ø	-Ø.77229	-Ø.83664
56	Ø.2571Ø	Ø.19Ø63	Ø.3537Ø
57	-18.16495	-17.4527Ø	-17.29149
58	4.49Ø58	6.4Ø985	6.96952
59	34.65996	41.25768	42.66293
CONSTANT	-2974.56787	-2976.1Ø693	-2955.851Ø7

Table F.1.d

Classification Matrix

Segment Before the Analysis	Predicted Segment Independent	Subordinate	Secondary	Total
Independent	54	3	0	57
Subordinate	2	116	5	123
Secondary	0	4	111	115
Previously Unclassified	20	65	48	133
Total Classified	76	188	164	428

Table F.1.e

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	57	26	45.6	8	14.0
Subordinate	123	75	60.9	4	3.3
Secondary	115	39	33.9	15	13.0
Total	295	140	47.6	27	9.2

* Chi^2 at .10 level with 14 degrees of freedom $\leq$ 7.78

** Chi^2 at .99 level with 14 degrees of freedom $\geq$ 29.12

Figure F.1

Two-Dimensional Representation

of the Labour Market Segments

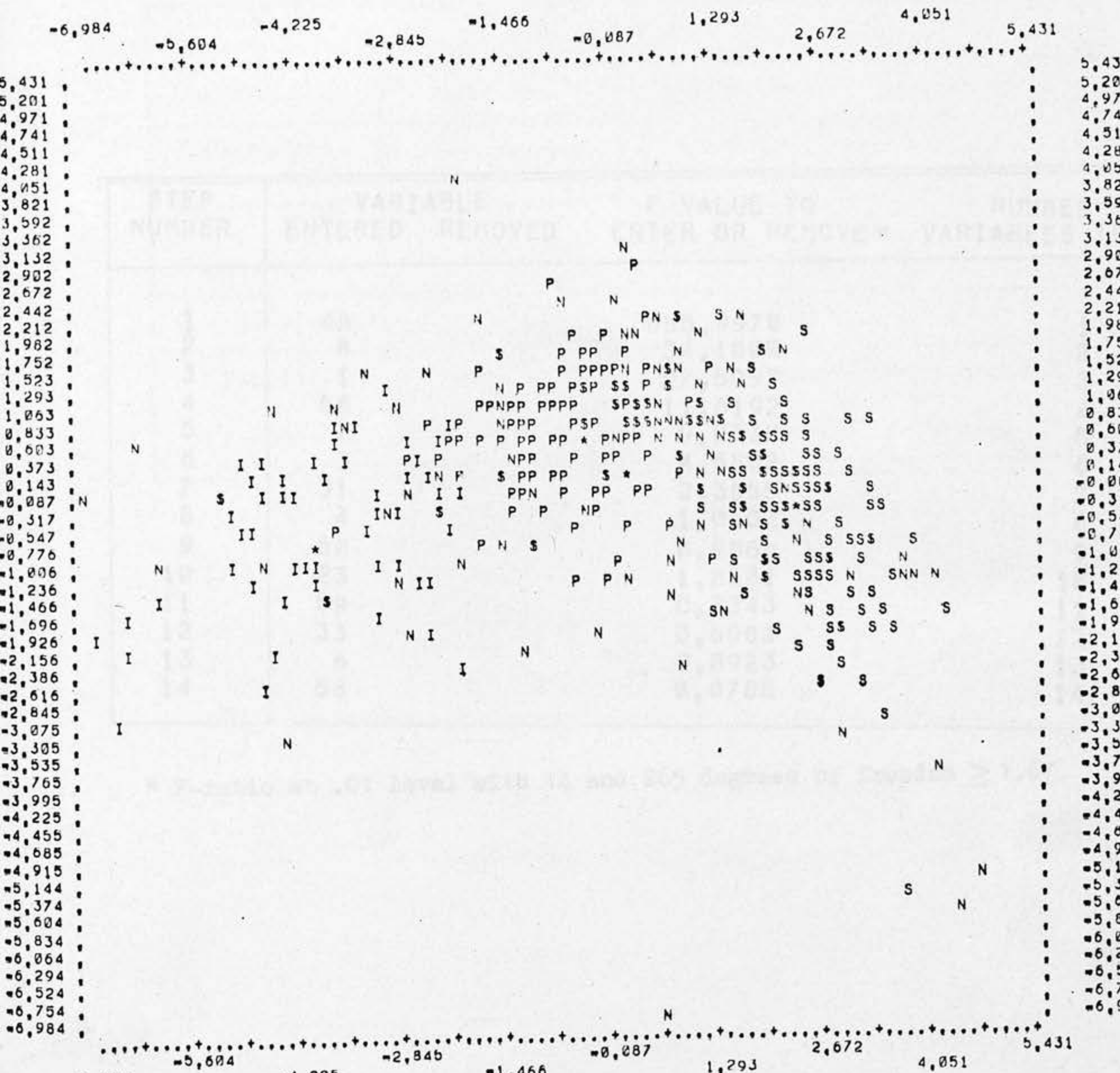


Table F.2.a

Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE *	NUMBER OF VARIABLES INCLUDED
1	48		655.9978	1
2	8		34.1000	2
3	1		27.5997	3
4	56		11.6192	4
5	7		9.0330	5
6	3		4.5882	6
7	31		2.3855	7
8	4		1.0102	8
9	60		0.9865	9
10	23		1.0681	10
11	59		0.7343	11
12	33		0.6983	12
13	6		0.0923	13
14	58		0.0788	14

* F-ratio at .01 level with 14 and 265 degrees of freedom ≥ 1.67

Table F.2.b

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	----------	43.477	121.991
Subordinate	43.477	----------	43.222
Secondary	121.991	43.222	----------
F - Statistic**	43.176		

* F-ratio at .01 level with 14 and 265 degrees of freedom 1.67

** F-ratio at .01 level with 28 and 530 degrees of freedom 1.72

Table F.2.c

Discriminant Functions

VARIABLE	FUNCTION INDPRI	PRISUB	SECSEC
1	0.55832	0.53128	0.43115
3	57.76736	58.13557	58.15709
4	56.95535	57.29620	57.50011
6	2.29317	2.23246	2.16623
7	3.09551	3.05487	2.86801
8	32.94612	31.26892	29.76324
23	-0.62931	-0.68831	-0.68582
31	0.29862	0.27056	0.24767
33	-0.29337	-0.29155	-0.27981
48	-0.42559	-0.61027	-0.69255
56	0.42383	0.35867	0.51276
58	28.12033	28.49339	28.78629
59	50.31059	56.50597	58.70880
60	16.27986	14.99954	15.04613
CONSTANT	-3200.33936	-3186.86475	-3152.38818

Table F.2.d

Classification Matrix

Segment Before the Analysis	Predicted Segment: Independent	Subordinate	Secondary	Total
Independent	53	1	0	54
Subordinate	0	115	1	116
Secondary	0	4	107	111
Previously Unclassified	22	67	58	147
Total Classified	75	187	166	428

Table F.2.e

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	54	22	40.7	8	14.8
Subordinate	116	74	63.8	5	4.3
Secondary	111	36	32.4	14	12.6
Total	281	132	47.0	27	9.6

* Chi^2 at .10 level with 14 degrees of freedom $\leq$ 7.78

** Chi^2 at .99 level with 14 degrees of freedom $\geq$ 29.12

Figure F.2

Two-Dimensional Representation

of the Labour Market Segments

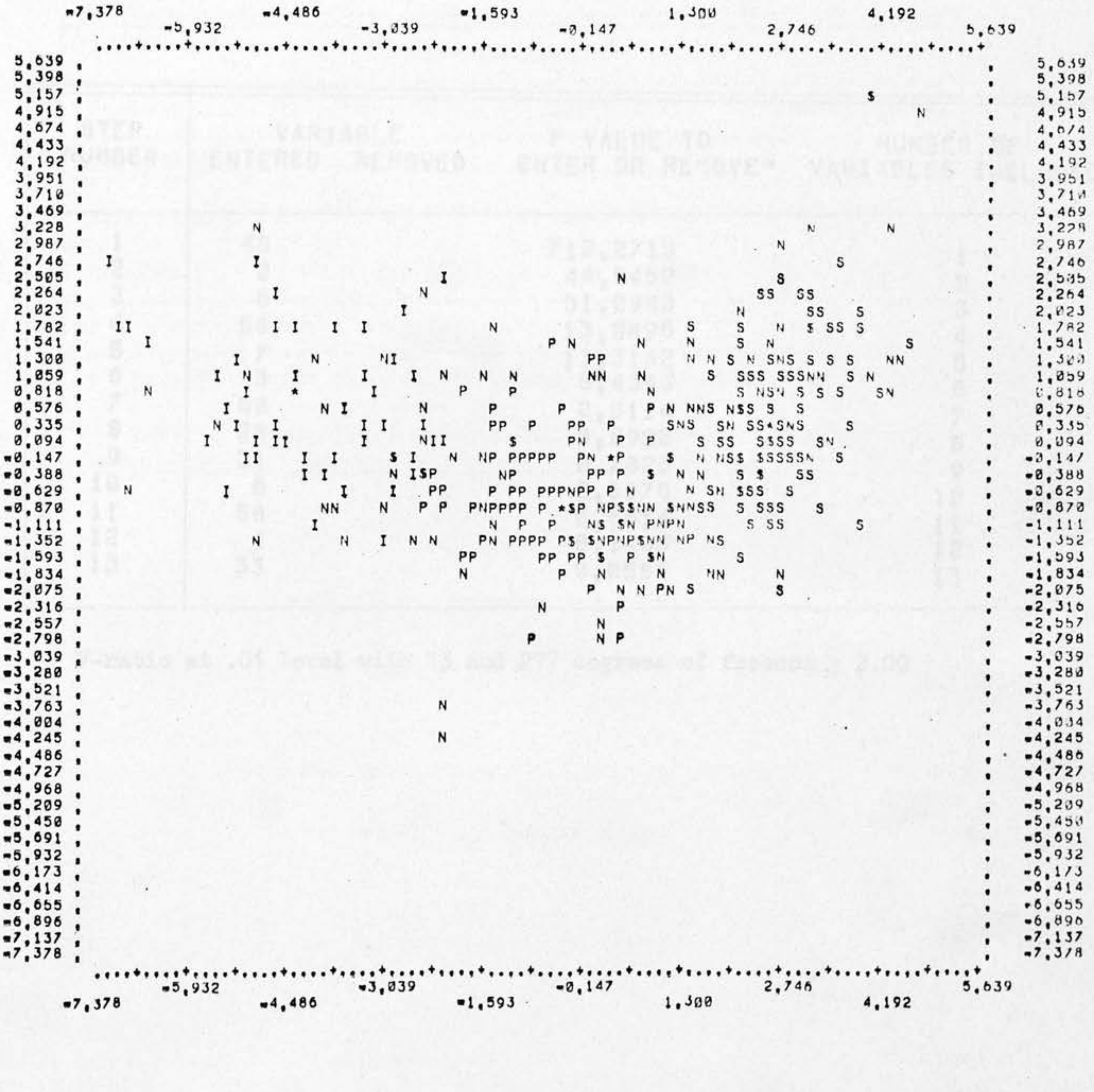


Table F.3.a

Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE*	NUMBER OF VARIABLES INCLUDED
1	48		712.2715	1
2	2		44.9452	2
3	8		31.0943	3
4	56		13.0496	4
5	7		11.3142	5
6	3		5.4343	6
7	60		2.0114	7
8	23		0.8920	8
9	31		0.7026	9
10	6		0.5978	10
11	58		0.6614	11
12	4		0.1406	12
13	33		0.0581	13

* F-ratio at .01 level with 13 and 277 degrees of freedom $\geq$ 2.00

Table F.3.b

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	----------	53.762	131.865
Subordinate	53.762	----------	55.242
Secondary	131.865	55.242	----------
F - Statistic**	52.500		

* F-ratio at .01 level with 13 and 277 degrees of freedom 2.00

** F-ratio at .01 level with 26 and 554 degrees of freedom 1.76

Table F.3.c

Discriminant Functions

VARIABLE	FUNCTION INDPRI	PRISUB	SECSEC
2	-0.31704	-0.30078	-0.16913
3	113.63113	113.85928	113.79472
4	113.64143	113.76949	113.90840
6	8.88161	8.95877	8.77529
7	4.08195	4.06231	3.83451
8	40.03685	38.19675	36.93916
23	0.47774	0.54203	0.65435
31	1.65282	1.61112	1.58571
33	-0.41592	-0.41805	-0.42222
48	-0.38848	-0.62813	-0.71832
56	0.82853	0.66684	0.87114
58	37.31552	36.84781	37.68613
60	24.54257	22.72052	24.23338
CONSTANT	-6045.60937	-6013.45703	-5983.39844

Table F.3.d

Classification Matrix

Segment Before the Analysis	Predicted Segment			
	Independent	Subordinate	Secondary	Total
Independent	47	0	0	47
Subordinate	2	136	2	140
Secondary	0	1	104	105
Previously Unclassified	28	50	58	136
Total Classified	77	187	164	428

Table F.3.e

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	47	17	36.2	6	12.8
Subordinate	140	75	53.6	9	6.4
Secondary	105	36	34.3	14	13.3
Total	292	128	43.8	29	9.9

* Chi^2 at .10 level with 13 degrees of freedom $\leq$ 7.05

** Chi^2 at .99 level with 13 degrees of freedom $\geq$ 27.69

Figure F.3

Two-Dimensional Representation

of the Labour Market Segments

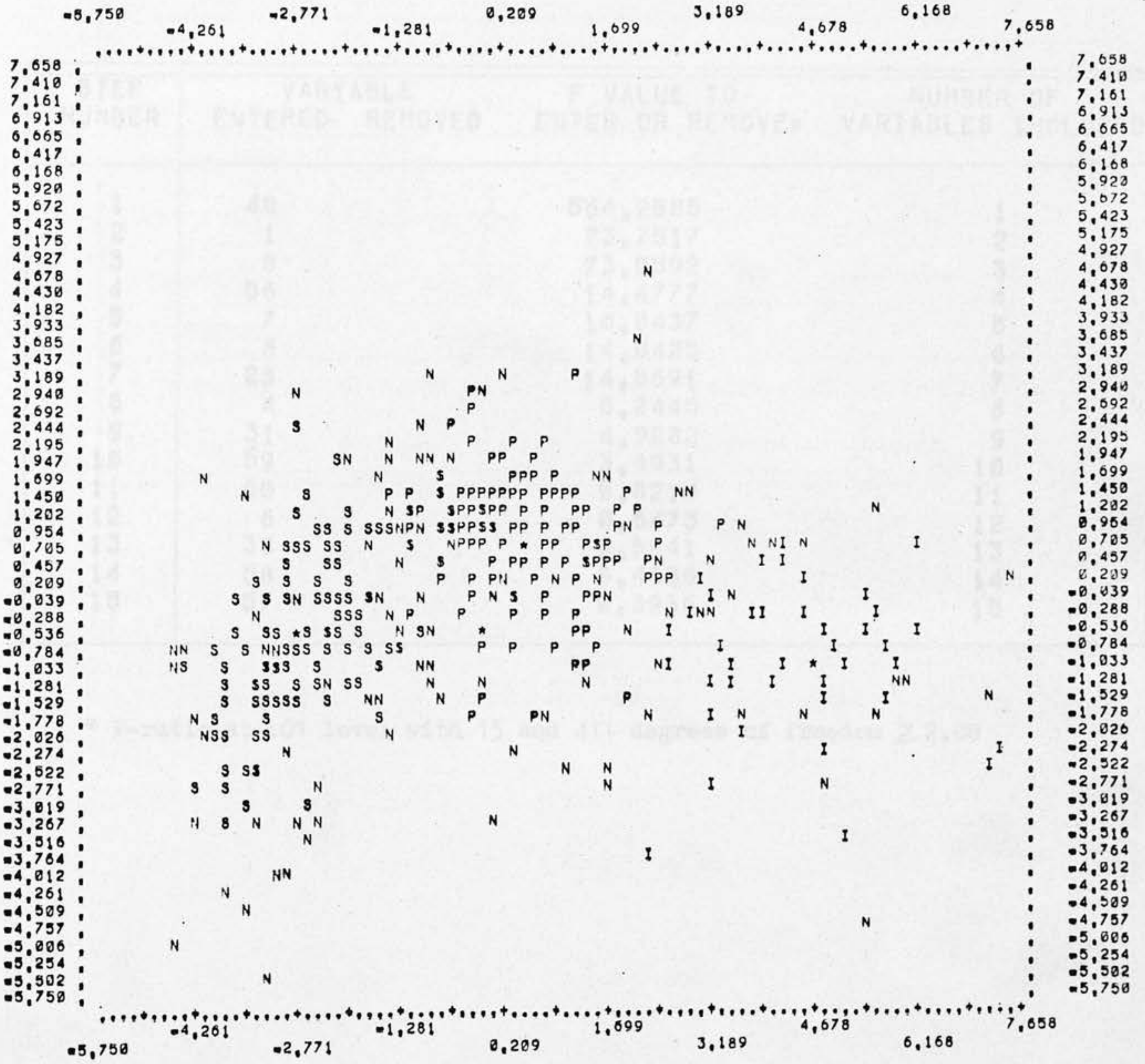


Table F.4.a

Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE*	NUMBER OF VARIABLES INCLUDED
1	48		564.2585	1
2	1		73.7517	2
3	8		73.0592	3
4	56		14.4777	4
5	7		16.0437	5
6	3		14.0425	6
7	23		14.8691	7
8	4		6.2445	8
9	31		4.9082	9
10	59		3.4931	10
11	60		0.8210	11
12	6		0.6473	12
13	33		0.5841	13
14	58		0.4836	14
15	57		0.3936	15

* F-ratio at .01 level with 15 and 411 degrees of freedom $\geq$ 2.08

Table F.4.b

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	------------	54.669	129.863
Subordinate	54.669	------------	52.369
Secondary	129.863	52.369	----------
F - Statistic**	56.917		
* F-ratio at .01 level with 15 and 411 degrees of freedom 2.08			
** F-ratio at .01 level with 30 and 822 degrees of freedom 1.72			

Table F.4.c

Discriminant Functions

VARIABLE	FUNCTION INDPRI	PRISUB	SECSEC
1	0,43419	0,40437	0,28587
3	44,23354	44,80371	44,96896
4	44,16396	44,67609	44,94133
6	4,54883	4,42781	4,31722
7	1,52799	1,50793	1,33173
8	21,53079	20,09492	19,10728
23	-0,26400	-0,17105	-0,09853
31	0,04064	-0,00593	-0,03396
33	-0,36254	-0,37229	-0,37072
48	-0,54033	-0,68630	-0,73166
56	0,86917	0,76010	0,85573
57	406,35596	410,84644	411,69336
58	425,69751	430,66772	432,05371
59	410,12695	406,24902	406,91016
60	430,63354	435,49805	437,44507
CONSTANT	-2589,81519	-2609,45581	-2597,96167

Table F.4.d

Classification Matrix

Segment Before the Analysis	Predicted Segment Independent	Subordinate	Secondary	Total
Independent	75	2	0	77
Subordinate	3	182	2	187
Secondary	0	10	154	164
Previously Unclassified	0	0	0	0
Total Classified	78	194	156	428

Table F.4.e

Significance of Classifications

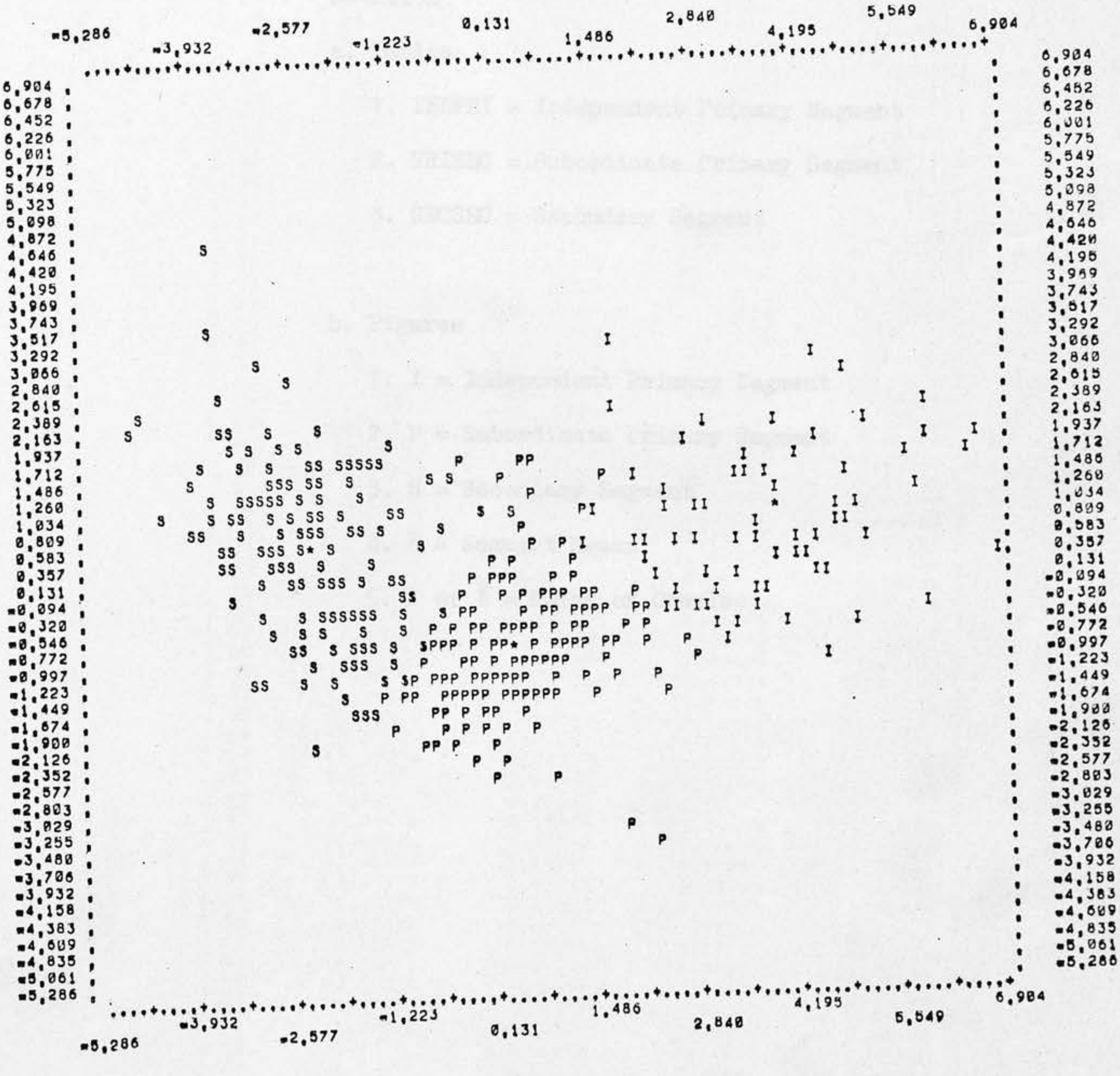
Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	77	36	46.8	12	15.6
Subordinate	187	133	71.1	7	3.7
Secondary	164	68	41.5	25	15.2
Total	428	147	34.3	44	10.3

* Chi^2 at .10 level with 15 degrees of freedom $\leq$ 8.55

** Chi^2 at .99 level with 15 degrees of freedom $\geq$ 30.60

Figure F.4

Two-Dimensional Representation of the Labour Market Segments



Appendix G

Empirical Approach: Hierarchical Classification on 295 Occupations

Notation

a. Tables

1. INDPRI = Independent Primary Segment
2. PRISEC = Subordinate Primary Segment
3. SECSEC = Secondary Segment

b. Figures

1. I = Independent Primary Segment
2. P = Subordinate Primary Segment
3. S = Secondary Segment
4. * = Segment Means
5. £ or $ = Point of Overlap

Table G.1.a

Means:

First Hierarchical Classification on 295 Occupations

VARIABLE	GROUP INDPRI	PRISUB	SECSEC	NEWGRP
1	61.81125	78.10252	55.04990	72.52061
2	38.18738	21.89651	44.94994	27.47871
3	97.71700	98.26028	92.97118	95.18526
4	1.54749	1.51273	6.57376	4.37341
5	0.73449	0.22615	0.45455	0.44062
6	0.57728	1.51066	0.72573	0.65326
7	40.31689	41.03267	39.35249	41.98137
8	15.91957	12.65879	12.89136	13.98241
9	53.86363	58.32498	48.31882	41.61345
10	70.10735	68.38327	50.09631	64.86215
11	9.10368	10.09800	14.72284	11.83631
12	20.78773	21.51770	35.18071	23.30069
13	69.78014	66.15353	70.33510	68.34865
14	19.62788	21.85281	17.56186	18.55678
15	46.51686	41.59540	45.60356	39.28368
16	23.53822	24.15897	30.02177	28.44946
17	35.29315	30.32532	37.35475	32.21657
18	27.34293	27.59851	30.10541	22.93988
19	37.36302	41.08566	32.53956	40.33162
20	81.49132	83.37317	64.24620	75.44434
21	8.50949	5.37513	11.12710	8.31794
22	9.99795	10.26045	24.62654	11.72570
23	1.88671	1.08620	7.66696	3.42486
24	0.04923	1.13836	0.01191	1.66638
25	1.42781	10.80140	4.95156	7.68355
26	21.32143	44.10608	6.30254	29.06721
27	6.15773	12.13908	4.76434	7.43903
28	12.10625	16.67468	11.91216	9.06619
29	10.94914	0.31980	2.30911	1.99312
30	3.17987	5.18973	3.61011	1.39755
31	1.99105	1.38244	18.50496	4.08984
32	0.55271	1.41306	3.83892	1.74840
33	31.25990	3.32811	23.78192	27.51318
34	9.11771	2.42044	12.34532	4.15852
35	0.0	0.0	0.0	0.75188
36	70.49278	85.88779	70.09932	74.73515
37	21.69435	6.84339	21.92371	15.98453
38	7.64105	7.10068	5.65601	8.86314
39	1.61541	3.18904	2.37483	3.40403
40	14.80520	16.20375	32.13026	17.06435
41	14.43331	16.95396	18.72170	17.99332
42	70.76013	66.84138	49.14784	62.68591
43	7.19433	5.51072	16.61528	6.53326
44	12.67374	11.54982	22.53853	15.19760
45	80.13065	81.94832	60.84590	73.75710
46	0.0	0.0	0.0	0.0
47	66.46303	66.13625	51.13164	68.74388
48	86.29424	58.73274	37.24725	64.91092
49	83.49474	82.92805	79.69943	81.87561
50	8.20568	6.67246	3.84458	7.30559
51	0.41152	0.34224	0.05968	0.70522
52	0.22645	0.46433	0.38619	1.12941
53	0.02345	0.07291	0.17115	0.35577
54	6.13071	10.66294	13.56857	12.25258
55	0.98835	2.07471	3.22932	1.37370
56	3.54863	6.13978	10.37973	6.66199
57	1.00000	0.03960	0.05660	0.42857
58	0.0	0.96040	0.26415	0.45865
59	0.0	0.0	0.07547	0.02256
60	0.0	0.0	0.60377	0.08271

Table G.1.b

Standard Deviations:

First Hierarchical Classification on 295 Occupations

VARIABLE	GROUP INDPRI	PRISUB	SECSEC	NEWGRP
1	31,77287	27,32973	38,07686	32,72456
2	31,77319	27,32991	38,07692	32,72459
3	2,58560	1,84197	5,93370	13,56830
4	2,07539	1,72130	5,91849	13,49689
5	1,78674	0,63713	1,15706	1,91654
6	0,96971	2,32976	1,43199	1,86025
7	5,60996	7,11367	9,94210	9,60688
8	1,88624	1,36521	1,23089	2,89757
9	26,21614	35,79210	32,18887	37,87317
10	13,22554	19,53896	22,81541	26,19598
11	6,16594	7,75417	12,60861	18,36852
12	12,22258	18,19661	24,08911	21,56543
13	10,71646	12,09279	15,28011	22,81863
14	9,52897	11,73192	11,98576	19,20465
15	11,82951	11,46683	12,01277	23,05939
16	8,89292	10,05440	12,64455	18,73030
17	15,05532	12,74065	17,63434	22,46248
18	12,96083	10,65357	11,31353	19,07329
19	14,10667	15,13946	15,31053	27,12395
20	11,49309	15,18337	21,08075	28,34450
21	8,07926	6,97588	10,01334	16,68074
22	9,36866	11,58578	20,60312	19,05728
23	10,19590	5,19083	26,63585	17,23671
24	0,28185	9,95321	0,06212	10,19151
25	4,58071	23,58479	16,86633	22,09044
26	27,84856	40,03635	13,53438	37,82443
27	17,67528	27,84895	15,52720	21,99683
28	24,47781	25,17963	22,01407	20,48937
29	23,75090	1,23563	7,66040	9,82644
30	5,11178	11,75925	9,64226	4,68973
31	9,96567	4,87565	33,83604	17,20583
32	2,46446	7,10467	14,98535	8,98028
33	39,39365	10,86308	32,61969	40,06900
34	22,43466	10,30439	28,71129	15,82140
35	0,0	0,0	0,0	8,67109
36	27,89706	16,80475	32,31445	30,00783
37	27,24092	12,54722	31,07343	27,18097
38	16,64699	11,64978	13,52360	18,91975
39	1,81520	3,74417	2,67306	10,44309
40	13,21195	13,19170	21,46745	21,74446
41	11,04327	11,02312	11,03886	20,19167
42	19,53133	18,77567	24,74863	27,94028
43	11,23528	11,82792	18,83382	16,42264
44	10,90201	10,26993	16,40823	19,71655
45	18,69565	18,89861	27,92204	29,01541
46	0,0	0,0	0,0	0,0
47	23,54205	21,81427	24,71765	24,49654
48	45,25478	25,96233	27,46182	36,10544
49	14,11949	11,61725	15,38661	22,69734
50	14,30214	10,46468	7,97197	13,39336
51	1,00648	1,06406	0,20866	4,69707
52	0,93057	1,83097	1,57109	7,89889
53	0,11240	0,35206	1,14437	2,47227
54	5,44926	8,78309	10,77163	16,37408
55	1,56489	2,30242	4,02609	2,04504
56	3,54597	5,59384	8,90248	9,13588
57	0,0	0,19600	0,23329	0,49674
58	0,0	0,19600	0,44510	0,50017
59	0,0	0,0	0,26668	0,14905
60	0,0	0,0	0,49379	0,27648

Table G.1.c
Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE*	NUMBER OF VARIABLES INCLUDED
1	57		1467.8840	1
2	58		216.5416	2
3	4		96.5756	3
4	6		19.3624	4
5	11		7.6507	5
6	22		6.3873	6
7	26		4.6886	7
8	55		4.2703	8
9	24		3.6814	9
10	43		3.3942	10
11	40		5.0964	11
12	32		4.1032	12
13	18		3.8647	13
14	34		3.3692	14
15	37		2.8981	15
16	12		2.3504	16
17	39		2.3153	17
18	25		2.9151	18
19	47		2.2530	19
20	17		2.2105	20
21	5		1.8963	21
22	13		1.4547	22
23	56		1.1666	23
24	9		1.1657	24
25	33		1.1043	25
26	44		1.1917	26
27	14		1.2141	27
28	15		1.5489	28
29	21		1.3953	29
30	51		1.2913	30
31	60		1.2293	31
32	27		0.9093	32
33	28		1.0718	33
34	54		0.7735	34
35	1		0.7203	35
36	7		0.5727	36
37	16		0.4872	37
38	30		0.4148	38
39	50		0.4052	39
40	36		2.0477	40
41	38		0.7862	41
42	53		0.7680	42
43	48		0.5800	43
44	49		0.4160	44
45	23		0.2800	45
46	31		0.2526	46
47	41		0.2334	47
48	19		0.1230	48
49	8		0.0631	49
50	52		0.0268	50

* F-ratio at .01 level with 50 and 243 degrees of freedom ≥ 1.60

Table G.1.d

Significance of Difference Between the Segments

	F - Matrix*		
Segment	Independent	Subordinate	Secondary
Independent	----------	69.915	60.529
Subordinate	69.915	----------	20.231
Secondary	60.529	20.231	----------
F - Statistic**	43.004		

* F-ratio at .01 level with 50 and 243 degrees of freedom 1.60

** F-ratio at .01 level with 100 and 486 degrees of freedom 1.41

Table G.1.e

Discriminant Functions

VARIABLE	FUNCTION INDPRI	PRISUB	SECSEC
1	0.17220	0.20967	0.17077
4	-2.62977	-1.76814	0.06855
5	-0.53398	-1.26837	-0.69661
6	1.76423	4.16600	3.67817
7	7.43233	7.23827	7.16261
8	32.80208	32.56786	32.45642
9	0.36008	0.38421	0.35371
11	0.04343	-0.17331	-0.34278
12	1.72138	1.74949	1.77968
13	2.39902	2.59437	2.62007
14	2.34639	2.51196	2.49377
15	0.35767	0.28202	0.22845
16	-0.15711	-0.16426	-0.20750
17	4.25039	4.13823	4.13270
18	3.95667	3.95734	4.01499
19	4.29632	4.29681	4.24739
21	-0.36266	-0.21420	-0.30906
22	0.21048	0.44254	0.55622
23	2.53159	2.46968	2.46740
24	1.03959	0.92380	0.78440
25	0.64635	0.73887	0.79342
26	0.49044	0.54377	0.50974
27	0.66603	0.72373	0.71852
28	0.30919	0.35772	0.34049
30	0.20104	0.16252	0.22037
31	0.14389	0.14534	0.11280
32	0.38326	0.77572	0.78812
33	-0.05920	-0.00762	0.01827
34	1.38906	1.55096	1.66476
36	2.53653	2.71646	2.75447
37	1.64072	1.76909	1.71044
38	3.43093	3.38190	3.62415
39	4.60459	4.23186	3.74344
40	-0.45170	-0.81452	-0.72926
41	0.16816	0.15588	0.19195
43	-0.28717	-0.21806	-0.31671
44	0.53173	0.39523	0.41822
47	-0.38872	-0.56884	-0.52229
48	-0.97119	-0.96010	-0.92254
49	1.51404	1.47930	1.42216
50	-0.11342	0.13069	-0.20151
51	2.50765	3.47134	2.76962
52	3.77628	3.89687	3.91778
53	2.75723	3.83347	2.64157
54	-1.77535	-1.64504	-1.56750
55	2.31425	1.82489	1.89402
56	1.71169	2.05090	1.89171
57	309.03174	245.16736	210.32986
58	295.08154	300.98633	262.62158
60	262.41553	260.93262	254.36661
CONSTANT	-1070.14771	-1050.72583	-1021.79883

Table G.1.f

Classification Matrix

Segment Before the Analysis	Predicted Segment Independent	Subordinate	Secondary	Total
Independent	141	0	0	141
Subordinate	3	98	0	101
Secondary	0	5	48	53
Previously Unclassified	57	54	22	133
Total Classified	201	157	70	428

Table G.1.g

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	141	77	54.6	14	9.9
Subordinate	101	56	55.4	19	18.8
Secondary	53	13	24.5	19	15.8
Total	295	146	49.5	52	17.6

* Chi^2 at .10 level with 50 degrees of freedom $\leq$ 37.50

** Chi^2 at .99 level with 50 degrees of freedom $\geq$ 76.10

Figure G.1

Two-Dimensional Representation
of the Labour Market Segments

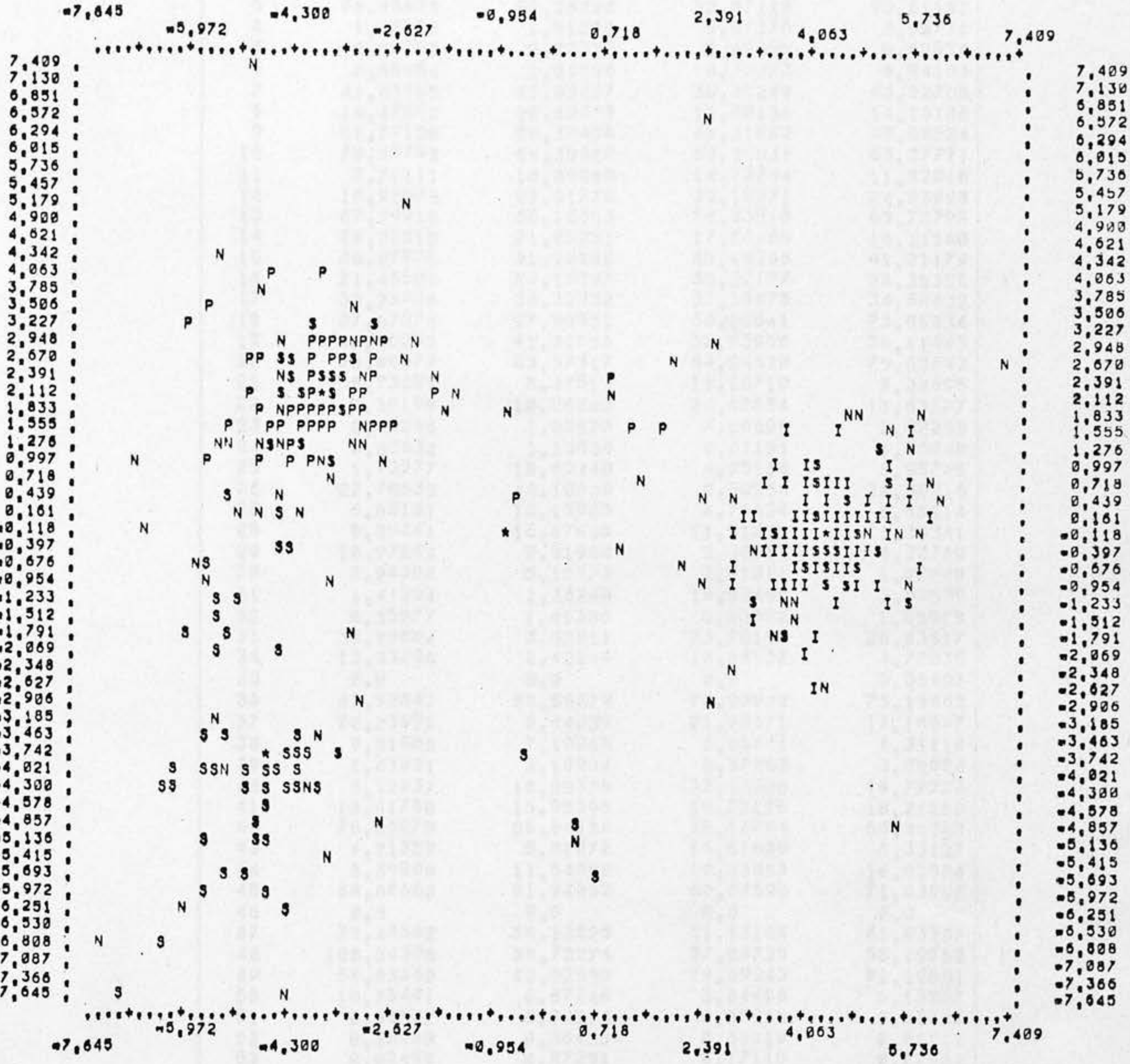


Table G.2:a

Means:
Second Hierarchical Classification on 295 Occupations

VARIABLE	GROUP INDPRI	PRISUB	SECSEC	NEWGRP
1	79.11682	78.10252	55.04990	60.37471
2	20.88225	21.89651	44.94994	39.62399
3	98.08693	98.26028	92.97118	95.61191
4	1.02478	1.51273	6.57376	3.95736
5	0.88763	0.22615	0.45455	0.42976
6	0.56454	1.51066	0.72573	0.64135
7	41.66765	41.03267	39.35249	40.82750
8	16.47083	12.65879	12.89136	14.16186
9	51.77158	58.32498	48.31882	45.80524
10	75.37793	68.38327	50.09631	63.27771
11	7.71111	10.09800	14.72284	11.92018
12	16.91000	21.51770	35.18071	24.80098
13	67.94910	66.15353	70.33510	69.70798
14	20.92610	21.85281	17.56186	18.11160
15	46.27975	41.59540	45.60356	41.21179
16	21.48509	24.15897	30.02177	28.35355
17	32.25836	30.32532	37.35475	34.64439
18	27.67876	27.59851	30.10541	23.85034
19	40.06265	41.08566	32.53958	38.11443
20	83.88377	83.37317	64.24620	75.63643
21	6.73329	5.37513	11.12710	9.33895
22	9.38199	10.26045	24.62654	11.63377
23	2.65296	1.08620	7.66696	2.62258
24	0.05632	1.13836	0.01191	1.26048
25	1.73977	10.80140	4.95156	5.95749
26	22.70535	44.10608	6.30254	26.38316
27	6.63101	12.13908	4.76434	6.86114
28	9.29341	16.67468	11.91216	11.36341
29	10.97064	0.31980	2.30911	4.20769
30	2.94304	5.18973	3.61011	1.97040
31	1.41264	1.38244	18.50496	3.88509
32	0.35577	1.41306	3.83892	1.55909
33	30.90604	3.32811	23.78192	28.63817
34	10.33286	2.42044	12.34532	4.72538
35	0.0	0.0	0.0	0.56497
36	67.72647	85.88779	70.09932	75.19662
37	22.23895	6.84339	21.92371	17.10547
38	9.91865	7.10068	5.65601	7.31118
39	1.43031	3.18904	2.37483	3.06084
40	8.92037	16.20375	32.13026	19.72778
41	12.41780	16.95396	18.72176	18.21289
42	78.66075	66.84138	49.14784	60.36340
43	4.21329	5.51072	16.61528	8.33127
44	8.89900	11.54982	22.53853	16.63884
45	86.88660	81.94832	60.84590	71.63907
46	0.0	0.0	0.0	0.0
47	79.48662	66.13625	51.13164	61.03967
48	108.24396	58.73274	37.24725	58.19762
49	85.63240	82.92805	79.69943	81.10661
50	10.75441	6.67246	3.84458	6.13257
51	0.53527	0.34224	0.05968	0.56439
52	0.30588	0.46433	0.38619	0.86142
53	0.02499	0.07291	0.17115	0.27232
54	5.79217	10.66294	13.56857	10.91629
55	0.63959	2.07471	3.22932	1.46903
56	2.86632	6.13978	10.37973	6.26197
57	1.00000	0.03960	0.05660	0.57062
58	0.0	0.96040	0.26415	0.34463
59	0.0	0.0	0.07547	0.01695
60	0.0	0.0	0.60377	0.06215

Table G.2.b

Standard Deviations:

Second Hierarchical Classification on 295 Occupations

VARIABLE	GROUP INDPRI	PRISUB	SECSEC	NEWGRP
1	15.94729	27.32973	38.07686	37.24287
2	15.94736	27.32991	38.07692	37.24269
3	2.60798	1.84197	5.93370	11.83191
4	1.70046	1.72130	5.91849	11.76932
5	2.08165	0.63713	1.15706	1.69989
6	1.08797	2.32976	1.43199	1.64252
7	5.34681	7.11367	9.94210	8.91553
8	1.87870	1.36521	1.23089	2.60059
9	29.34979	35.79210	32.18867	34.61661
10	9.55420	19.53896	22.81541	23.71901
11	5.90401	7.75417	12.60861	16.15169
12	9.76644	18.19661	24.08911	19.89705
13	11.40947	12.09279	15.28011	20.26219
14	10.46536	11.73192	11.98576	16.93306
15	13.15904	11.46683	12.01277	20.66074
16	9.07008	10.05440	12.64455	16.54498
17	15.05754	12.74065	17.63434	20.89745
18	13.24203	10.65357	11.31353	17.69577
19	14.79198	15.13946	15.31053	24.34755
20	11.01237	15.18337	21.08075	25.12941
21	6.91363	6.97588	10.01334	15.23592
22	9.92362	11.58578	20.60312	16.96594
23	12.22058	5.19083	26.63585	14.99936
24	0.32510	9.95321	0.06212	8.85472
25	5.39580	23.58479	16.86633	19.38254
26	29.91632	40.03635	13.53438	34.93236
27	19.95895	27.84895	15.52720	19.86803
28	22.62898	25.17963	22.01407	22.67636
29	26.15759	1.23563	7.66040	12.76091
30	5.26765	11.75925	9.64226	4.80040
31	9.78079	4.87565	33.83604	15.76029
32	1.28687	7.10467	14.98535	8.01143
33	39.78760	10.86308	32.61969	39.73303
34	25.88251	10.30439	28.71129	14.86839
35	0.0	0.0	0.0	7.51646
36	29.59300	16.80475	32.31445	28.35248
37	28.93199	12.54722	31.07343	26.28880
38	19.34662	11.64978	13.52360	16.81985
39	1.90644	3.74417	2.67306	9.09558
40	6.59407	13.19170	21.46745	20.74002
41	10.48561	11.02312	11.03886	18.32448
42	13.65878	18.77567	24.74863	26.33136
43	7.31879	11.82792	18.83382	16.35861
44	8.48516	10.26993	16.40823	18.11012
45	12.97724	18.89861	27.92204	27.39755
46	0.0	0.0	0.0	0.0
47	11.11035	21.81427	24.71765	26.54182
48	36.18945	25.96233	27.46182	34.32999
49	15.49755	11.61725	15.38661	20.19119
50	16.26167	10.46468	7.97197	12.07132
51	1.13689	1.06406	0.20866	4.08422
52	1.11008	1.83097	1.57109	6.85698
53	0.11975	0.35206	1.14437	2.14650
54	4.83603	8.78309	10.77163	14.73676
55	1.22608	2.30242	4.02609	2.01926
56	3.28610	5.59384	8.90248	8.14703
57	0.0	0.19600	0.23329	0.49639
58	0.0	0.19600	0.44510	0.47659
59	0.0	0.0	0.26668	0.12945
60	0.0	0.0	0.49379	0.24211

Table G.2.c
Summary Table

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE*	NUMBER OF VARIABLES INCLUDED
1	57		1008.0381	1
2	58		183.7994	2
3	4		98.8158	3
4	6		16.7071	4
5	11		6.4975	5
6	22		6.7909	6
7	26		3.8530	7
8	43		4.1672	8
9	55		4.5731	9
10	24		3.6890	10
11	40		3.6211	11
12	18		3.5003	12
13	47		3.3878	13
14	32		4.6544	14
15	34		3.2185	15
16	9		2.3749	16
17	39		2.6048	17
18	25		3.4860	18
19	5		1.8703	19
20	15		1.8546	20
21	37		1.6298	21
22	17		1.2415	22
23	21		1.6969	23
24	51		1.5069	24
25	13		1.3272	25
26	14		1.3679	26
27	56		1.3381	27
28	44		1.7653	28
29	28		1.0725	29
30	33		0.8546	30
31	27		1.7610	31
32	50		0.9374	32
33	38		4.2128	33
34	30		1.3518	34
35	36		1.1092	35
36	19		0.6817	36
37	2		0.5914	37
38	60		0.5891	38
39	53		0.8642	39
40	16		0.4917	40
41	29		0.3967	41
42	48		0.3449	42
43	41		0.3037	43
44	8		0.2877	44
45	23		0.3175	45
46	7		0.2739	46
47	54		0.3244	47
48	49		0.1381	48
49	10		0.1130	49

* F-ratio at .01 level with 49 and 200 degrees of freedom ≥ 1.62

Table G.2.d

Significance of Difference Between the Segments

F - Matrix*			
Segment	Independent	Subordinate	Secondary
Independent	------------	52.767	51.592
Subordinate	52.767	------------	20.593
Secondary	51.592	20.593	----------
F - Statistic**	36.647		

* F-ratio at .01 level with 49 and 200 degrees of freedom 1.62

** F-ratio at .01 level with 98 and 400 degrees of freedom 1.42

Table G.2.e
Discriminant Functions

VARIABLE	FUNCTION INDPRI	PRISUB	SECSEC
2	0.12711	0.11352	0.17453
4	-1.69044	-0.77833	1.21347
5	0.78956	0.06649	0.67491
6	-0.00996	2.12645	1.72867
7	5.97975	5.79383	5.65627
8	33.09062	32.83574	32.15324
9	0.28149	0.29391	0.24344
10	-0.41551	-0.38459	-0.36417
11	-0.46277	-0.66331	-0.83176
13	1.91789	2.09813	2.12991
14	1.48480	1.62351	1.62745
15	0.43853	0.36468	0.25800
16	-0.03313	-0.07546	-0.12256
17	2.92204	2.79334	2.77920
18	3.09430	3.04937	3.10309
19	3.74132	3.72483	3.61696
21	-0.40971	-0.20415	-0.33987
22	0.93476	1.16400	1.25048
23	2.86593	2.78555	2.81373
24	1.28073	1.17604	1.07007
25	1.20962	1.30882	1.39987
26	1.19846	1.24990	1.25580
27	1.52045	1.58345	1.62680
28	1.04583	1.09884	1.11443
29	1.05310	1.05219	1.08956
30	0.77686	0.74798	0.89408
32	1.19891	1.59189	1.64187
33	0.60732	0.64996	0.74334
34	2.31177	2.47104	2.62452
36	2.01224	2.21099	2.17698
37	0.82340	0.95991	0.84619
38	3.43044	3.39928	3.68765
39	5.31421	5.03736	4.39411
40	0.25478	-0.09405	0.11827
41	0.48381	0.46317	0.51058
43	-0.65549	-0.53820	-0.68571
44	0.06831	-0.06146	-0.00088
47	-0.19456	-0.37318	-0.31272
48	-0.86263	-0.88417	-0.85072
49	2.42743	2.37513	2.35583
50	0.02282	0.27946	-0.16169
51	1.48184	2.54603	1.33577
53	-1.64448	-0.57329	-1.80584
54	-0.79268	-0.74788	-0.67208
55	1.65039	1.07161	1.03601
56	2.05238	2.32857	2.15617
57	266.82471	205.44202	173.96318
58	259.61304	260.24268	223.76117
60	239.10051	232.82132	225.54970
CONSTANT	-990.14868	-959.99438	-925.93945

Table G.2.f

Classification Matrix

Segment Before the Analysis	Predicted Segment Independent	Subordinate	Secondary	Total
Independent	97	0	0	97
Subordinate	3	98	0	101
Secondary	0	5	48	53
Previously Unclassified	100	54	23	177
Total Classified	200	157	71	428

Table G.2.g

Significance of Classifications

Segment	Number of Occupations	Significant Core Members*	Per Cent	Significant Non-Members**	Per Cent
Independent	97	48	49.5	6	6.2
Subordinate	101	59	58.4	15	14.8
Secondary	53	13	24.5	18	33.9
Total	251	120	47.8	39	15.5

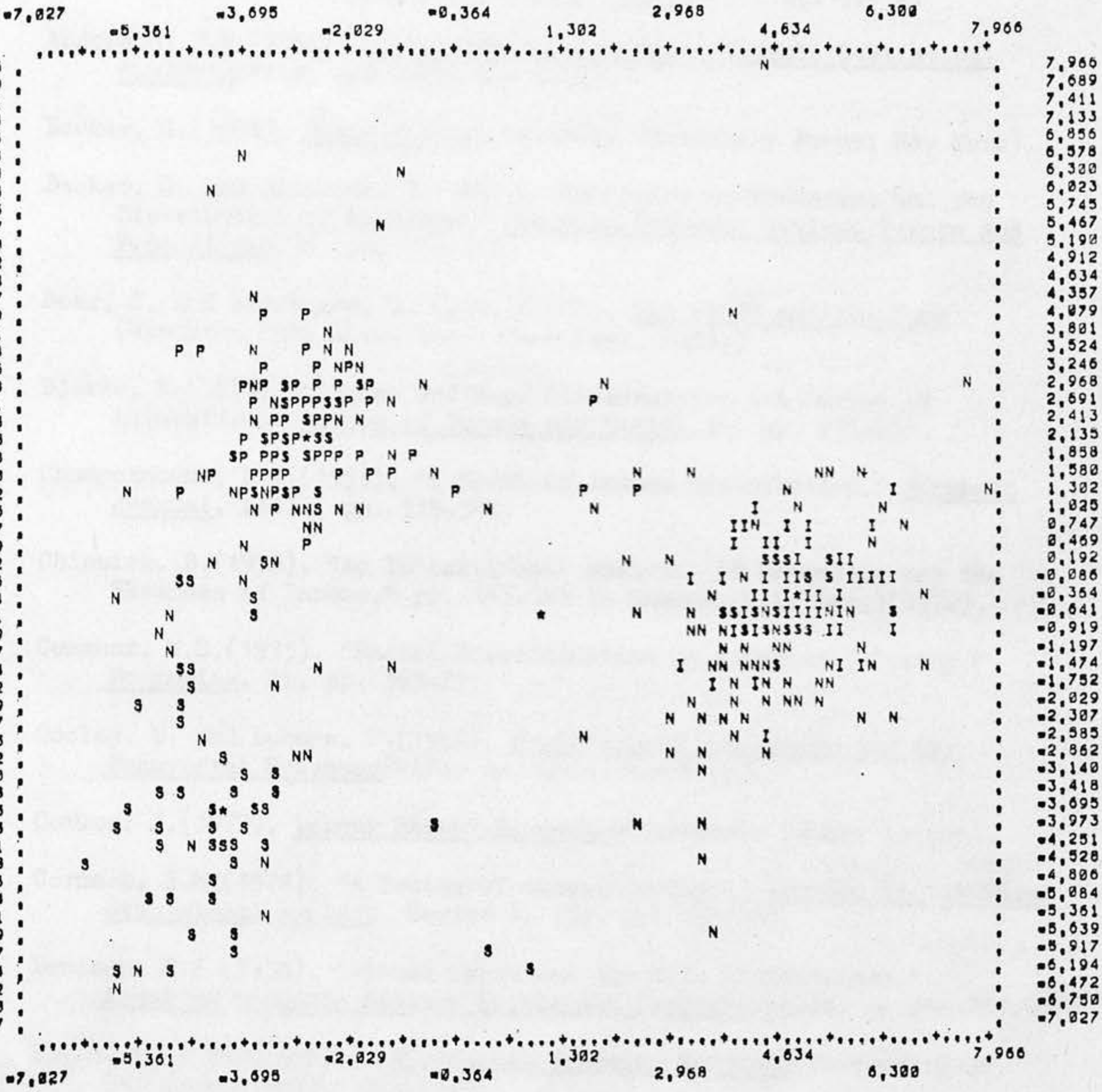
* Chi^2 at .10 level with 49 degrees of freedom $\leq$ 36.70

** Chi^2 at .99 level with 49 degrees of freedom $\geq$ 75.64

Figure G.2

Two-Dimensional Representation

of Labour Market Segments



References

Adams, F.G.(1958), "The Size of Individual Incomes: Socio-Economic Variables and Chance Variation," Review of Economics and Statistics, 40,pp.390-397.

Aitchison, J. and Brown, J.(1969), The Lognormal Distribution (Cambridge University Press; Cambridge, England).

Al-Samarrie, A. and Miller, H.P.(1967), "State Differentials in Income Concentration," American Economic Review, 57, pp. 59-72.

Anderson, T.W.(1958), An Introduction to Multi-Variate Statistical Analysis(Wiley and Sons; New York).

Becker, G.(1964), Human Capital(Columbia University Press; New York).

Becker, G. and Chiswick, B.(1966), "Economics of Education and the Distribution of Earnings," American Economic Review; Papers and Proceedings,56, pp. 358-369.

Beer, S. and Barringer, R. (eds.)(1970), The State and the Poor (Winthrop Publishers Inc.; Cambridge, Mass.).

Bjerke, K.(1970), "Income and Wage Distributions - A Survey of Literature," Review of Income and Wealth,16, pp. 235-253.

Champernowne, D.G.(1953), "A Model of Income Distribution," Economic Journal, LXIII, pp. 318-351.

Chiswick, B.(1970), "An Interregional Analysis of Schoolong and the Skewness of Income," pp. 157-183 in Hansen, W.L. (ed.)(1970).

Comanor, W.S.(1973), "Racial Discrimination in American Industry," Economica, XL, pp. 363-378.

Cooley, W. and Lohnes, P.(1962), Multivariate Procedures for the Behavioral Sciences(Wiley and Sons; New York).

Corina, J.(1972), Labour Market Economics(Heinemann Books; London).

Cormack, R.M.(1971), "A Review of Classification," Journal of the Royal Statistical Society, Series A, 134, pp. 321-353.

Denison, E.F.(1954), "Income Types and the Size Distribution," American Economic Review: Papers and Proceedings,44, pp.254-269.

Dixon, W.J. (ed.)(1971), Biomedical Computer Programs(University of California Press; Berkeley).

Doeringer, P.(1969), Programs to Employ the Disadvantaged(Prentice Hall; New Jersey).

Doeringer, P.(1971), Internal Labour Markets, Technology, and Labour Force Adjustment(D.C. Heath; Lexington, Mass.).

Doeringer, P. and Piore, M.(1971), Internal Labour Markets and Manpower Analysis(D.C. Heath; Lexington, Mass.).

Dunlop, J.(1950), Wage Determination Under Trade Unions(A. Kelley; New York).

Dunlop, J.(1957), "Contemporary Wage Theory," pp. 3-33 in Dunlop, J. (ed.)(1957).

Dunlop, J. (ed.)(1957), The Theory of Wage Determination(MacMillan and Co.; London).

Farbman, M.(9173), "Income Concentration in the Southern United States," Review of Economics and Statistics,55, pp. 333-340.

Ferguson, C.E.(1972), Microeconomic Theory(3rd. ed., Richard Irwin Inc.; Homewood, Illinois).

Friedman, M.(1953), "Choice, Chance, and the Personal Distribution of Income," Journal of Political Economy,61, pp. 277-290.

Garvy, G.(1954), "Functional and Size Distribution of Income and Their Meaning," American Economic Review,44, pp. 236-253.

Gendarme, R.(1968), "Reflections on Approaches to the Problems of Distribution in Underdeveloped Countries," pp. 361-388 in Marchal, J. and Ducros, B. (eds.)(1968).

Ginsberg, M.(1929), "Interchange between Social Classes," Economic Journal,39, pp. 554-565.

Gramm, W.L.(9173), "The Labour Force Decision of Married Female Teachers: A Discriminant Analysis Approach," Review of Economics and Statistics, 55, pp.341-348.

Hansen, W.L. (ed.)(1970), Education, Income, and Human Capital(Columbia University Press; New York).

Hanushek, E.(1973), "Regional Differences in the Stucture of Earnings," Review of Economics and Statistics,55, pp. 204-214.

Hunter, L.C. and Robertson, D.J.(1969), Economics of Wages and Labour (MacMillan Press; London).

Jackson, J.M.(1970), Wages and Labour Economics(McGraw Hill; London).

Johnston, J.(1972a), Econometric Methods(2nd. ed., McGraw Hill; London).

Johnston, J.(1972b), "A Model of Wage Determination under Bilateral Monopoly," Economic Journal,82, pp. 837-853.

Keat, P.G.(1960), "Long Run Trends in the Occupation Wage Structure 1900-1956," Journal of Political Economy,68(6), pp. 584-600.

Kendall, M.G.(1966), "Discrimination and Classification," pp. 165-187 Krishnaiah, P.R. (ed.)(1966).

Kendall, M. and Stuart, A.(1966), The Advanced Theory of Statistics: Volume III(C. Griffin and Co.; London).

Kerr, C.(1957a), "Labour's Income Share and the Labour Movement," pp. 260-299 in Taylor, G. and Pierson, F. (eds.)(1957).

Kerr, C.(1957b), "Wage Relationships - Comparative Impact of Market and Power Forces," pp. 173-193 in Dunlop, J. (ed.)(1957).

Klein, L.(1962), Introduction to Econometrics(Prentice Hall; New Jersey).

Kravis, I.B.(1960), "International Differences in Income Distribution," Review of Economics and Statistics,42, pp. 408-416.

Krishnaiah, P.R. (ed.)(1966), Multivariate Analysis(Academic Press; New York).

Lampman, R.(1954), "Changes in Income Inequality," American Economic Review,54, pp. 251-268.

Lydall, H.(1968), The Structure of Earnings(Clarendon Press; Oxford).

Maher, J.(1961), "Index of Wage Rates for Selected Indutries, 1946-1957," Review of Economics and Statistics,43, pp. 277-282.

Marchal, J.(1957), "Wage Theory and Social Groups," in Dunlop, J. (ed.) (1957).

Marchal, J. and Ducros, B. (eds.)(1968), Distribution of National Income (MacMillan Press; London).

McCormick, B.J.(1969), Wages(Penguin Books; Middlesex).

McCormick, B.J. and Smith, E.O. (eds.)(1968), The Labour Market (Penguin Books; Middlesex).

Miller, H.P.(1955), Income of the American People(Wiley & Sons; New York).

Mincer, J.(1958), "Investment in Human Capital and Personal Income," Journal of Political Economy,LXVI, pp. 281-301.

Pen, J.(1971), Income Distribution(Allan Lane The Penguin Press; London).

Phelps Brown, E.H. and Hart, P.E.(1952), "Share of Wages in National Income," Economic Journal,LXII, pp. 253-274.

Phillips, J.D.(1960), "Labour's Share and Wage Parity," Review of Economics and Statistics,42, pp. 164-174.

Piore, M.(1970), "Jobs and Training," pp. 53-83 in Beer, S. and Barringer, R. (eds.)(1970).

Piore, M.(1972), "Notes for a Theory of Labour Market Stratification," MIT Department of Economics Working Paper No. 95.

Rao, C.R.(1970), Advanced Statistical Methods in Biometric Research (Hafner Publishing Co.; Daven, Conn.).

Reder, M.W.(1955), "The Theory of Occupational Wage Differentials," American Economic Review,44, pp. 833-852, reprinted in McCormick, B.J. and Smith, E.O. (eds.)(1968), pp. 203-227.

Reder, M.W.(1969), "Partial Survey of Theory of Income Size Distribution," pp. 205-255 in Soltow, L. (ed.)(1969).

Ross, A.M.(1948), Trade Union Wage Policy(University of California Press; Los Angeles).

Rulon, P., Tiedeman, D., Tatswoka, M., and Langmuir, C.(1967), Multivariate Statistics for Personnel Classification(Wiley & Sons; New York).

Schultz, T.P.(1969), "Secular Trends and Cyclical Behaviour of Income Distribution in the United States: 1944-1965," pp. 75-101 in Soltow, L. (ed.)(1969).

Soltow, L.(1960), "Distribution of Income Related to Changes in Distribution of Education, Occupation, Age," Review of Economics and Statistics,42, pp. 450-453.

Soltow, L. (ed.)(1969), Six Papers on the Size Distribution of Wealth and Income(Columbia University Press; New York).

Taylor, G.W. and Pierson, F.C. (eds.)(1957), New Concepts in Wage Determination(McGraw Hill; New York).

Tinbergen, J.(1970), "A Positive and a Normative Theory of Income Distribution," Review of Income and Wealth,16, pp. 221-235.

Tinbergen, J.(1972), "Impact of Education on Income Distribution," Review of Income and Wealth,18, pp. 255-265.

Turner, A.(1933), "Theory of Industrial Disputes," Review of Economic Studies,1, pp. 154-156.

Turner, H.A.(1957), "Inflation and Wage Differentials in Great Britain," pp. 123-135 in Dunlop, J. (ed.)(1957), reprinted in McCormick, B,J. and Smith, E.O. (eds.)(1968), pp. 228-242.

U.S. Department of Commerce(1971), Public Use Samples of Basic Records from the 1960 and 1970 Censuses(U.S. Government Printing Office; Washington, D.C.).

U.S. Department of Commerce(1972a), 1970 Occupation and Industry Classification Systems in Terms of Their 1960 Occupation and Industry Elements(U.S. Government Printing Office; Washington, D.C.).

U.S. Department of Commerce(1972b), Public Use Samples of Basic Records from the 1970 Census: Description and Technical Documentation (U.S. Government Printing Office; Washington, D.C.).

U.S. Department of Commerce(1973), Occupational Characteristics of the of the 1970 Census of the Population(U.S. Government Printing Office; Washington, D.C.).

Ward, J.(1963), "Hierarchical Grouping to Optimize an Objective Function," Journal of the American Statistical Association,58, pp. 236-244.

Whitaker, J.(1974), "Marshallian System 1881: Distribution and Growth," Economic Journal,84, pp. 1-18.

Wishart, D. (1969a), "An Algorithm for Hierarchical Classification," Biometrics,25, pp. 165-170.

Wishart, D.(1969b), Clustan(St. Andrews University; St. Andrews).